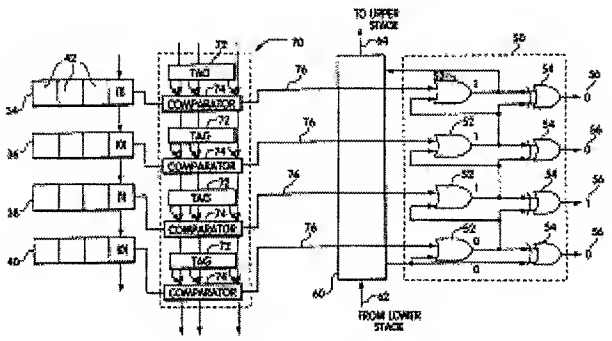
Tag-based scheduling system for digital communication switch

Publication number: US5455825
Publication date: 1995-10-03
Inventor: LAUER HUGH (US); SHEN CHIA (US); GHOSH ABHIJIT (US)
Applicant: MITSUBISHI ELECTRIC RESEARCH L (US)
Classification:
- international: ***H04Q3/00; H04L12/56; H04Q3/52;*** **H04Q3/00; H04L12/56; H04Q3/52;** (IPC1-7): H04L12/56
- European: H04L12/56E3
Application number: US19940234385 19940428
Priority number(s): US19940234385 19940428

Also published as:
JP8056230 (A)

Abstract of **US5455825**

A switch for digital communication networks includes a queuing system capable of implementing a broad class of scheduling algorithms for many different applications and purposes, with the queueing system including means for providing numerical tags to incoming cells or packets, the values of the tags being calculated when incoming cells or packets arrive at the switch. A queue and search module is provided to select cells or packets for transmission based on these tags. The combination of the tags and the queue and search module enables simple and fast implementations of a wide variety of scheduling algorithms, including algorithms for supporting communication traffic with real time requirements, continuous media such as audio and video, and traffic requiring very fast response. Furthermore, multiple classes of traffic are supported in a single network switch, each class having its own scheduling algorithm and policy. The queue and search module is designed for VLSI implementation, and in one embodiment supports an ATM switch with 16 ports, each port operating at 622 megabits per second.



(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平8－56230

(43)公開日　平成８年(1996)２月27日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｈ０４Ｌ　12/28				
12/56				
Ｈ０４Ｑ　3/00				
		9466－5K	Ｈ０４Ｌ　11/ 20	Ｈ
		9466－5K		Ｇ

審査請求　未請求　請求項の数22　ＯＬ　（全 22 頁）　最終頁に続く

(21)出願番号　特願平7－103908

(22)出願日　平成７年(1995)４月27日

(31)優先権主張番号　０８／２３４３８５

(32)優先日　1994年４月28日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号

(72)発明者　フー　ラウア
アメリカ合衆国　マサチューセッツ州　コンコルド　ブローダーロード　69

(72)発明者　チア　シェン
アメリカ合衆国　マサチューセッツ州　ソマービル　モリソンアベニュー　169

(72)発明者　アブヒジット　ゴーシュ
アメリカ合衆国　カリフォルニア州　バークレイ　スラターレーン　42

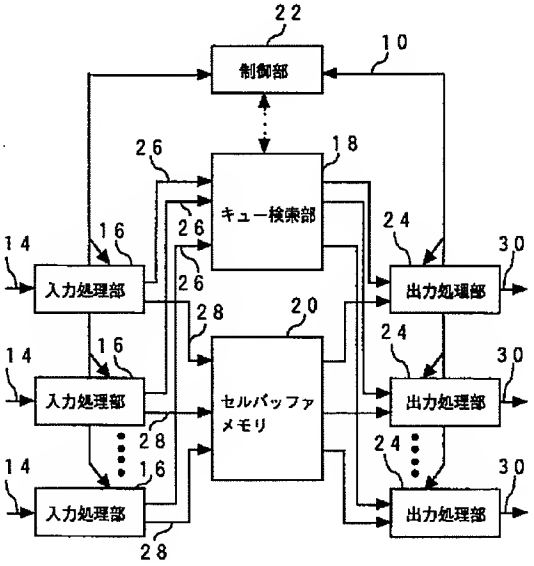
(74)代理人　弁理士　高田　守　（外４名）

(54)【発明の名称】　スイッチングシステム

(57)【要約】

【目的】　ダイナミック・スケジューリング・アルゴリズムをサポートできるタグベースのスイッチングシステムを得る。

【構成】　入力したセルに宛先ビットと優先度を示すタグを付加し、タグの値により出力順を決定する。タグの値の比較は、木構造で配置された比較器により並列に行うので、処理速度の速いスイッチングシステムが実現できる。



1

【特許請求の範囲】

【請求項１】　複数の入力リンクと複数の出力リンクを持ち、各入力リンクによりネットワーク内の１以上の宛先を示す宛先情報を含むヘッダフィールドとデータとを格納したセルを受信するスイッチと、

上記ヘッダフィールドの宛先情報を、各ビットが各出力リンクに対応しておりそのビットが有意状態である時そのセルが最終宛先に到着するようにそのビットに対応する出力リンクに対してそのセルを出力することを示している宛先ビットのベクトルに変換する変換手段と、

各セルに対して、上記宛先ビットのベクトルを付加するベクトル付加手段と、

上記ベクトルを付加されたセルを１つ以上の出力リンクに接続する接続手段とを備え、

上記接続手段は、上記ベクトルを付加されたセルを到着順に格納するキューイング手段と、キューイング手段に格納されたセルの中から到着順とは異なる順に各出力リンクに対して各出力リンクに対応する宛先ビットが有意状態のセルを選択する選択手段とを備え、

上記選択手段は、セルの出力をスケジュールするスケジュール手段を有し、上記スケジュール手段は、選択順を決定するアルゴリズムを記憶する手段と、

スイッチの動作に先だってそのアルゴリズムを設定する手段と、

スイッチの動作中にそのアルゴリズムを再設定する手段と、

上記各出力リンクに対してセル送信時間内に上記アルゴリズムを実行する手段を有しており、

上記スイッチは、更に、上記出力リンクに選択したセルを出力する出力手段を備えたことを特徴とするディジタル通信ネットワーク用のスイッチングシステム。

【請求項２】　上記出力手段は、上記ベクトルが除去されたセルを出力することを特徴とする請求項１記載のスイッチングシステム。

【請求項３】　上記アルゴリズムは、ダイナミック・プライオリティ・スケジューリング・アルゴリズムと、スタティック・プライオリティ・スケジューリング・アルゴリズムと、ラウンド・ロビン・アルゴリズムと、これらの組み合わせのいずれかから選択されることを特徴とする請求項１記載のスイッチングシステム。

【請求項４】　上記スケジュール手段は、各出力リンクにセルを出力する順番を決定するために設定されたスケジューリングアルゴリズムに基づいて、キューイング手段の中に格納されたセルの出力順序を示す値をタグ値として計算し、このタグ値を有するタグをセルに付加するタグ付加手段を有し、上記選択手段は、タグ値と宛先ビットの値に基づいて、キューイング手段の中に格納されたセルをサーチするサーチ手段とを備え、選択手段は、各出力リンクに対応する宛先ビットの値が有意状態であるセルの中で最も小さいタグ値を持つセルをサーチして

2

選択し、もし、最も小さいタグ値を持つセルが２以上存在する場合に、先に到着したセルを選択することを特徴とする請求項１記載のスイッチングシステム。

【請求項５】　上記キューイング手段は、先頭から末尾に至る複数のレジスタからなるＦＩＦＯ回路を備え、各レジスタは、出力を待っているセルを示すものであり、各レジスタは、上記宛先ビットのベクトルとタグ値のバイナリーコードを保持し、各レジスタは、先頭から末尾の方向でセルの到着順を示していることを特徴とする請求項４記載のスイッチングシステム。

【請求項６】　上記サーチ手段は、上記レジスタの数と同じ数の複数の比較回路を備え、各比較回路は、各レジスタに対応して設けられ、各比較回路は、対応したレジスタのタグ値と末尾側にある次のレジスタに対応した比較回路からの出力値を比較するようにリニアに配置され、各比較回路は、対応するレジスタのタグ値が末尾側にある次に比較回路から出力されたタグ値より小さい場合であって、かつ、選択した出力リンクに対する宛先ビットが有意状態である場合に、出力ビットをオンにすることにより、上記比較回路は、選択した出力リンクに対して最小のタグ値を伝搬させ、待ち行列の中からセルを選択することを特徴とする請求項５記載のスイッチングシステム。

【請求項７】　上記サーチ手段は、上記レジスタの数の半分に当たる複数の比較回路と追加の比較回路を備え、上記比較回路を階層的に配置し、ルートと枝と葉を持つツリーを構成し、葉と枝のサブセットによりサブツリーを構成し、葉に配置された各比較回路は、２つの隣り合うレジスタのタグ値を比較して小さいタグ値を有するレジスタを識別し、上記追加の比較回路は、２つのサブツリーからのタグ値を比較し、選択した出力リンクに対して、最小のタグ値を持つレジスタが存在しているサブツリーを識別し、上記ツリーは、選択した出力リンクに対して、最小のタグ値を伝搬することを特徴とする請求項５記載のスイッチングシステム。

【請求項８】　複数の出力リンクと出力リンクを示すアドレスを含んだセルを受信する複数の入力リンクと、

上記アドレスを宛先ビットに変換する変換手段と、

上記入力リンクのセルを少なくとも１つの出力リンクに接続する接続手段とを備え、

上記接続手段は、入力リンクから出力リンクへセルの出力をスケジュールするスケジュール手段を備え、

スケジュール手段は、

各出力リンクにセルを出力する順番を決定するために、予め設定されたスケジューリングアルゴリズムに基づいて、どのセルを出力すべきかを示す値を生成し、生成した値をタグ値として有するタグをセルに付加するタグ手段と、

共通キューと、

上記共通キューに、各セルに対して宛先ビットとタグ値

を挿入する手段と、
上記宛先ビットとタグ値に基づいて、上記共通キューから、対応している宛先の中から最小のタグ値を持つセルをサーチするサーチ手段と、
宛先に対応した出力リンクに対して最小のタグ値を持つセルを出力する出力手段とを備えたことを特徴とするディジタル通信ネットワーク用のスイッチングシステム。

【請求項９】　上記サーチ手段は、最小のタグ値をシリアルに評価していくリニアサーチ手段を有していることを特徴とする請求項８記載のスイッチングシステム。

【請求項１０】　上記サーチ手段は、木構造を用いて同時に複数対のセルを評価していくロガリズミックサーチ手段を有していることを特徴とする請求項８記載のスイッチングシステム。

【請求項１１】　上記サーチ手段は、リニアサーチ手段とロガリズミックサーチ手段を結合したサーチ手段を有していることを特徴とする請求項８記載のスイッチングシステム。

【請求項１２】　上記タグ手段は、
先頭から末尾までのキューを形成するように配置された一連のタグレジスタと、
複数の比較器と、
上記タグレジスタと比較器を接続する接続手段と、
上記セルを対応するタグレジスタにより識別できるように記憶する記憶手段と、
記憶されたセルのタグ値をタグレジスタに設定する設定手段とを備え、
上記比較器は、タグレジスタのタグ値と１つ前の比較器からの出力値を比較して小さい方の値を決定するとともに、
小さい方の値と宛先ビットを出力する出力手段と、
小さい方の値と宛先ビットに基づいて、記憶されたセルを識別する識別手段とを備えたことを特徴とする請求項８記載のスイッチングシステム。

【請求項１３】　上記比較器は、一連のビットから構成された出力を持ち、この一連のビットの中に、宛先ビットが有意状態であり、同じ宛先ビットが有意状態であるキュー中の他のセルのタグ値よりも小さいタグ値を持つことを示す特定ビットを有しており、
各比較器は、各タグ値とキューの末尾方向にあるタグの最小値とを比較することにより、宛先別に各セルのプライオリティを設定することを特徴とする請求項１２記載のスイッチングシステム。

【請求項１４】　上記比較器は、２つのタグ値を比較する手段と、小さい方の値を出力する手段とを有し、１つのタグ値に対して１つの追加の入力ビットを備え、上記比較器の出力は、比較対象となるタグ値を示す追加の入力ビットにより条件付けられていることを特徴とする請求項１２記載のスイッチングシステム。

【請求項１５】　上記追加の入力ビットは、宛先ビットが有意状態になっているセルのタグ値を特定するものであることを特徴とする請求項１４記載のスイッチングシステム。

【請求項１６】　上記スイッチングシステムは、更に、最小のタグ値を持つセルが２以上存在する場合に、キューの先頭に近いセルを選択する手段を備えたことを特徴とする請求項１２記載のスイッチングシステム。

【請求項１７】　上記タグ手段は、枝を経由してルートに連結された葉を持った階層的木構造を有し、各枝は、一対のタグレジスタと隣り合うタグレジスタの間に設けられたタグレジスタ用比較器と、２つのタグレジスタ用比較器からの出力を入力する追加の比較器を有しており、上記追加の比較器は、宛先ビットが有意状態になっており、小さい方のタグ値を持つタグレジスタに対応するセルを識別する手段を有していることを特徴とする請求項８記載のスイッチングシステム。

【請求項１８】　上記タグレジスタ用比較器は、小さい方のタグ値と２つの比較ビットを出力し、２つの比較ビットの内、一方の比較ビットは一方のセルが小さい方のタグ値を持っていることを示し、２つの比較ビットの内、他方の比較ビットは他方のセルが小さい方のタグ値を持っていることを示すものであり、上記選択手段は、小さい方のタグ値を持っているセルを識別するために、宛先ビットと比較ビットを入力する複数のＡＮＤゲートを有し、複数のＡＮＤゲートは、ルートにおいてただ１つのＡＮＤゲートが有意な出力を有するように構成されており、このＡＮＤゲートの有意な出力により出力リンクに出力されるセルを示すことを特徴とする請求項１７記載のスイッチングシステム。

【請求項１９】　上記ＡＮＤゲートは、そのＡＮＤゲートに対応したセルが適切な宛先ビットを有しているかを判定するために用いられるものであり、上記タグレジスタ用比較器は、異なるＡＮＤゲートへ接続された出力ビットを有し、このタグレジスタ用比較器からの出力ビットは、対応するセルの宛先ビットとＡＮＤを取られ、その結果の出力が有意状態である場合は、対応するセルが宛先ビットが有意状態であり、最小のタグ値を持つセルであることを示すことを特徴とする請求項１８記載のスイッチングシステム。

【請求項２０】　上記スイッチングシステムは、更に、２以上のセルが最小のタグ値を有する場合、先着のセルを識別する手段を有することを特徴とする請求項１９記載のスイッチングシステム。

【請求項２１】　上記タグレジスタ用比較器は、先頭と末尾を持つキューを定義し、上記追加の比較器は、２入力から小さい値を出力するとともに、キューの先頭方向にある２入力の値がキューの末尾方向にある２入力の値以下である場合に、有意状態を出力することを特徴とする請求項１７記載のスイッチングシステム。

【請求項２２】　上記スイッチングシステムは、更に、

5

上記木構造のルート方向への上位層レベルにおいて、下位層レベルのどの枝が小さい方のタグ値を有しているか決定することにより小さい方のタグ値を有する枝を識別する識別手段を有し、上記識別手段は、各タグに対応してそのタグの宛先ビットとそのタグが最小のタグ値を有することを示すビットとのＡＮＤを取るためのＡＮＤゲートを有していることを特徴とする請求項１７記載のスイッチングシステム。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は、セルやパケットがノード間でネットワークを転送されるディジタル通信ネットワークに関するものである。さらに、この発明は、ネットワーク中でセルが、あるノードから次のノードへ転送される順序をスケジューリングするための改良された方式を備えたディジタル通信ネットワークに関する。

【０００２】

【従来の技術】一般に、ディジタル通信ネットワークでは、メッセージや情報の流れがパケットやセルと呼ばれる連続した小単位に細分化され、このセルやパケットが、ノードからノードへ伝送される。また、各ノードにおいて、スイッチがセルやパケットを伝送する順序と伝送する次のノードとを選択する。その結果、ディジタル情報はその最終宛先へ適時に到着する。これらのネットワークスイッチは、様々な特性のネットワーク通信をサポートできるのが望ましい。例えば、適時な到着をリアルタイムに厳しく保証することを要求する通信、音声やビデオ用の連続的なメディア通信、早急な返答を要求する通信が含まれる。

【０００３】重要なディジタル通信ネットワークの１つとして、ＡＴＭ（Ａｓｙｎｃｈｒｏｎｏｕｓ　Ｔｒａｎｓｆｅｒ　Ｍｏｄｅ：非同期伝送モード）ネットワークがある。ＡＴＭネットワークは、ネットワーク中のあるポイント（ノード）から、あるほかのポイント（ノード）へデータを伝送するのに用いられる。ここで、データや情報は、連続した固定サイズの小さなセルに細分化されて、ネットワークのノード間で伝送される。ここで述べているノードは、ネットワークの複数のノード間でパケットやセルを高速にスイッチングしたり、ルーティングしたりするＡＴＭスイッチを含む。ＡＴＭネットワークの一般的な原理は、Ｊ・ブライアン・リールズとダニエル・Ｃ・スウィンハートによる、「発生するギガビット環境とローカルＡＴＭの役割」（ＩＥＥＥコミュニケーションマガジン、３０巻、４類、１９９２年４月、５２－５８頁）、及びＣ・ラムによる、「ＡＴＭ方式への高速化」（ユニックス　レビュー、１０巻、１０類、１９９２年１０月、２９－３６頁）に述べられている。

【０００４】ＡＴＭネットワークの重要な課題の１つは、各スイッチによって伝送されるセルを、スケジューリングすることである。セルは通常、各スイッチの待ち

6

行列内に、バッファリングされる。セルの渋滞（輻輳）がないものとすると、これらのセルは、スイッチで入力リンクから受信され、直ちに出力リンクをとおして別の宛先へ送信される。しかし、セルが複数の入力リンクを通して到着し、同じ出力リンクへ同時に送信されなければならない場合、セルが望ましい順序で伝送されるには、セルの待ち行列（キュー）をつくることが必要である。

【０００５】どのセルがどの時間にどの順序で伝送されるかのスケジューリングを調整するためには、ＦＩＦＯ（Ｆｉｒｓｔ　Ｉｎ　Ｆｉｒｓｔ　Ｏｕｔ：先入れ先出し）方式による順序付けを使うのが普通である。ＦＩＦＯ方式においては、スイッチに、先に到着するセルが、次に到着するセルに優先して伝送される。リアルタイムアプリケーションをサポートするネットワークの場合、セルに優先度がつけられ、セルの優先度によって、セルは別々の待ち行列に記憶される。その後、セルは、その別個の待ち行列の優先度に従って、指示された順序で伝送される。これらの単純な方式は、限られた数と種類のリアルタイムアプリケーションのみをサポートしうる。なぜならば、ＦＩＦＯ方式と優先度スケジューリング機能は、データ損失のない適時な伝送に対して、限られた保証のみを提供できるからである。例えば、ＡＴＭローカル・エリア・ネットワーク、あるいは、ＡＴＭワイド・エリア・ネットワークの初期世代のスイッチは、単に、入力ネットワークリンクから出力ネットワークリンクへセルを転送するというスケジューリングをしたりディスパッチしたりという、大変簡易なスケジューリングアルゴリズムを備えている。通信は、ＡＴＭ法則によるＦＩＦＯ順序付けで処理されるが、少数の、通常２種類の優先順位が、リアルタイム通信要求を持つアプリケーションをサポートするために、かなり限定された方法で提供されている。静的に割り当てられた優先度は、近年の進んだ変化に富む多種多様なアプリケーションを持つ中位サイズのローカル・エリア・ネットワークに対してさえも、ほとんど十分でない。さらに、アプリケーションが音声やビデオなど連続的なメディアを伝送する必要がある場合、もしくは、アプリケーションが実時間で期待した返答を必要とする場合、ネットワークサービスの質や適時性に関して予測や保証をすることは実質的に不可能である。

【０００６】過去２０年にわたって、実時間やマルチメディア・コンピューティング及び通信に関して十分な研究がなされてきた。また、多くのスケジューリングアルゴリズムが細部にわたり発明され、研究されてきた。例えば、Ｊ・ジャンゴク・バーとＴ・スーダによる「ＡＴＭネットワークにおける通信量制御スキームとプロトコルの分析」（ＩＥＥＥ会報７９巻、第2版、１９９１年２月、１７０－１８９頁）は、ＡＴＭスイッチにおいて一般的なスケジューリング機能によって、実現されなけ

ればならない多数の通信コントロールスキームを説明している。さらに、Ｈ・ジャングらは、「転送速度に基づくサービス原則の比較」（Ｈ・ジャングとＳ・ケシャブ、ＡＣＭ　ＳＩＧＣＯＭＭ会報、’９１、チューリッヒ、１９９１年９月）、「レート制御静的優先度待ち」（Ｈ・ジャングとＤ・フェラーリ、国際コンピュータサイエンス学会技術報告＃ＴＲ－９２－００３、カリフォルニア、バークレー）の論文で、転送速度に基づくサービス原則と転送速度により制御された静的優先度キューイングを説明している。スケジューリングアルゴリズムのさらなる例は、以下の論文、論説に記載されている。Ｗ・Ａ・ホーン、「簡易なスケジューリングアルゴリズム」（ナーバル　リサーチ　ロジスティッククオータリー、２１巻、１９７４年、１７７－１８５頁）、Ｊ．Ｒ．ジャクソン「最大遅延量を最小にするプロダクションラインのスケジューリング」（マネージメント　サイエンス　リサーチ　プロジェクト、調査報告４３，ＵＣＬＡ、１９５５年１月）、Ｃ・Ｒ・カルマネク，Ｈ・カナキア，Ｓ・ケシャブ、「高速ネットワークのためのレート制御サーバー」（ＩＥＥＥグローバル　テレコミュニケーション　コンファレンス、カリフォルニア、サンディエゴ、１９９０年１２月、３００．３．１－３００．３．９頁）、Ａ・クマー，Ｊ・パレク「統合サービスネットワークにおけるフロー制御への汎用プロセッサーのシェアリングアプローチ」（ｐｈＤ論文、ＭＩＴ、１９９２年２月）、Ｈ・Ｔ・カング，Ａ・チャップマン、「ＡＴＭネットワークに対するＦＣＶＣ（Ｆｌｏｗ－Ｃｏｎｔｒｏｌｌｅｄ　Ｖｉｒｔｕａｌ　Ｃｈａｎｎｅｌｓ：フロー制御バーチャルチャネル）提案」（ネットワークプロトコルに関する１９９３年国際会議会報、カリフォルニア、サンフランシスコ、１９９３年、１０月１９－２２日）、Ｃ・Ｌ・リュウとＪ・Ｗ・レイランド、「ハードの実時間環境におけるマルチプログラミングのためのスケジューリングアルゴリズム」（ＡＣＭジャーナル、２０巻、第１版、１９７３年）、Ｌ・ジャング、「バーチャルクロック：パケットスイッチネットワークのための新しい通信量コントロールアルゴリズム」（ＡＣＭ　ＳＩＧＣＯＭＭ会報、ペンシルバニア、フィラデルフィア、１９９０年９月、１９－２９頁）、Ｌ・ジャング、「バーチャルクロック：パケットスイッチネットワークのための新しい通信量コントロールアルゴリズム」（コンピューターシステムにおけるＡＣＭトランザクション、９巻、第２版、１９９１年５月、１０１－１２４頁）、Ｑ・ジェング，Ｋ・シン、「２地点間パケットスイッチネットワークにおける実時間チャネルの確立能力について」（コミュニケーションにおけるＩＥＥＥトランザクション、１９９４年３月）。

【０００７】これらのアルゴリズムの多くは、ＡＴＭネットワークにおける実行には、不適当とされてきたことに注意しなければならない。なぜならば、こうしたアルゴリズムは、一般的な環境には、専門的すぎるからである。また、キューのサーチに時間がかかりすぎるからである。

【０００８】しかし、ＡＴＭスケジューリングのためのＶＬＳＩシーケンサーチップが、Ｈ・ジョナサン・カオとネクデット・アズンによる論文、「ＡＴＭ通信シェイパーと行列管理のためのＶＬＳＩシーケンサーチップ」（ＩＥＥＥソリッドステイトサーキットジャーナル、２７巻、第１１版、１９９２年１１月）において述べられている。このシステムにおいて、セルは出力ポートにスイッチングされる。各ポートは、それ自体の待ち行列とシーケンサーチップを含んでいる。各ポートにおいて、セルは、シーケンサーチップによって優先順位に分類される。このシステムは、改良された優先度スケジューリングを実現しているが、スイッチの各出力ポートに対して別々の分類回路を要求する。その結果、コストが増加し、スイッチのデザインの柔軟性が減少する。また、性能低下の可能性もある。

【０００９】共用バッファスケジューリング方式を用いたＡＴＭスイッチに関する重要な提案が、Ｈ・近藤ほか，Ｋ・大島ほかによる以下の２つの論文で述べられている。Ｈ・近藤，Ｈ・山中，Ｍ・石脇，Ｙ・松田，Ｍ・中谷、「ＡＴＭスイッチＬＳＩのための効率的なセルフタイムの行列アーキテクチャー」（カスタム　インテグレーテド　サーキット　コンファレンス、サンディエゴ、１９９４年５月）、Ｋ・大島，Ｈ・山中，Ｈ・斉藤，Ｈ・山田，Ｓ・小浜，Ｈ・近藤，Ｙ・松田、「ＳＴＳタイプの共用バッファリングとそのＬＳＩ実現に基づく新しいＡＴＭスイッチアーキテクチャー」（国際スイッチングシンポジウム’９２会報、日本、横浜、１９９２年１０月、３５９－３６３頁）。このＡＴＭスイッチデザインの重要な要素は、共通の待ち合わせ機能と、統計的な多重化を発展させるため共用バッファと、より迅速な性能とより低いコストである。このスイッチは、出力バッファリングを用いるほかのＡＴＭスイッチとは、すべての入力ポートからの入力セルが直接共通のバッファメモリに蓄積されるという点で、区別される。出力は、ＦＩＦＯ方式や単純な優先度方式に基いて選択される。このようなスイッチは、単純な優先度で満足するような、長距離テレコミュニケーションに対しては十分に機能するが、このスイッチには、ファクトリーオートメーションや、パワープラントコントロール、フルモーションビデオなどリアルタイムアプリケーションをサポートする待ち行列制御の機能はない。

【００１０】

【発明が解決しようとする課題】この発明は、以上のような課題を解決するためになされたもので、スタティックなスケジューリングアルゴリズムだけでなく、ダイナミックなスケジューリングアルゴリズムに則った待ち行列制御が可能なスイッチングシステムを得ることを目的

としている。

【００１１】

【課題を解決するための手段】この発明に係るスイッチングシステムは、複数の入力リンクと複数の出力リンクを持ち、各入力リンクによりネットワーク内の１以上の宛先を示す宛先情報を含むヘッダフィールドとデータとを格納したセルを受信するスイッチと、上記ヘッダフィールドの宛先情報を、各ビットが各出力リンクに対応しておりそのビットが有意状態である時そのセルが最終宛先に到着するようにそのビットに対応する出力リンクに対してそのセルを出力することを示している宛先ビットのベクトルに変換する変換手段と、各セルに対して、上記宛先ビットのベクトルを付加するベクトル付加手段と、上記ベクトルを付加されたセルを１つ以上の出力リンクに接続する接続手段とを備え、上記接続手段は、上記ベクトルを付加されたセルを到着順に格納するキューイング手段と、キューイング手段に格納されたセルの中から到着順とは異なる順に各出力リンクに対して各出力リンクに対応する宛先ビットが有意状態のセルを選択する選択手段とを備え、上記選択手段は、セルの出力をスケジュールするスケジュール手段を有し、上記スケジュール手段は、選択順を決定するアルゴリズムを記憶する手段と、スイッチの動作に先だってそのアルゴリズムを設定する手段と、スイッチの動作中にそのアルゴリズムを再設定する手段と、上記各出力リンクに対してセル送信時間内に上記アルゴリズムを実行する手段を有しており、上記スイッチは、更に、上記出力リンクに選択したセルを出力する出力手段を備えたことを特徴とする。

【００１２】上記出力手段は、上記ベクトルが除去されたセルを出力することを特徴とする。

【００１３】上記アルゴリズムは、ダイナミック・プライオリティ・スケジューリング・アルゴリズムと、スタティック・プライオリティ・スケジューリング・アルゴリズムと、ラウンド・ロビン・アルゴリズムと、これらの組み合わせのいずれかから選択されることを特徴とする。

【００１４】上記スケジュール手段は、各出力リンクにセルを出力する順番を決定するために設定されたスケジューリングアルゴリズムに基づいて、キューイング手段の中に格納されたセルの出力順序を示す値をタグ値として計算し、このタグ値を有するタグをセルに付加するタグ付加手段を有し、上記選択手段は、タグ値と宛先ビットの値に基づいて、キューイング手段の中に格納されたセルをサーチするサーチ手段とを備え、選択手段は、各出力リンクに対応する宛先ビットの値が有意状態であるセルの中で最も小さいタグ値を持つセルをサーチして選択し、もし、最も小さいタグ値を持つセルが２以上存在する場合に、先に到着したセルを選択することを特徴とする。

【００１５】上記キューイング手段は、先頭から末尾に至る複数のレジスタからなるＦＩＦＯ回路を備え、各レジスタは、出力を待っているセルを示すものであり、各レジスタは、上記宛先ビットのベクトルとタグ値のバイナリーコードを保持し、各レジスタは、先頭から末尾の方向でセルの到着順を示していることを特徴とする。

【００１６】上記サーチ手段は、上記レジスタの数と同じ数の複数の比較回路を備え、各比較回路は、各レジスタに対応して設けられ、各比較回路は、対応したレジスタのタグ値と末尾側にある次のレジスタに対応した比較回路からの出力値を比較するようにリニアに配置され、各比較回路は、対応するレジスタのタグ値が末尾側にある次に比較回路から出力されたタグ値より小さい場合であって、かつ、選択した出力リンクに対する宛先ビットが有意状態である場合に、出力ビットをオンにすることにより、上記比較回路は、選択した出力リンクに対して最小のタグ値を伝搬させ、待ち行列の中からセルを選択することを特徴とする。

【００１７】上記サーチ手段は、上記レジスタの数の半分に当たる複数の比較回路と追加の比較回路を備え、上記比較回路を階層的に配置し、ルートと枝と葉を持つツリーを構成し、葉と枝のサブセットによりサブツリーを構成し、葉に配置された各比較回路は、２つの隣り合うレジスタのタグ値を比較して小さいタグ値を有するレジスタを識別し、上記追加の比較回路は、２つのサブツリーからのタグ値を比較し、選択した出力リンクに対して、最小のタグ値を持つレジスタが存在しているサブツリーを識別し、上記ツリーは、選択した出力リンクに対して、最小のタグ値を伝搬することを特徴とする。

【００１８】また、この発明に係るスイッチングシステムは、複数の出力リンクと出力リンクを示すアドレスを含んだセルを受信する複数の入力リンクと、上記アドレスを宛先ビットに変換する変換手段と、上記入力リンクのセルを少なくとも１つの出力リンクに接続する接続手段とを備え、上記接続手段は、入力リンクから出力リンクへセルの出力をスケジュールするスケジュール手段を備え、スケジュール手段は、各出力リンクにセルを出力する順番を決定するために、予め設定されたスケジューリングアルゴリズムに基づいて、どのセルを出力すべきかを示す値を生成し、生成した値をタグ値として有するタグをセルに付加するタグ手段と、共通キューと、上記共通キューに、各セルに対して宛先ビットとタグ値を挿入する手段と、上記宛先ビットとタグ値に基づいて、上記共通キューから、対応している宛先の中から最小のタグ値を持つセルをサーチするサーチ手段と、宛先に対応した出力リンクに対して最小のタグ値を持つセルを出力する出力手段とを備えたことを特徴とする。

【００１９】上記サーチ手段は、最小のタグ値をシリアルに評価していくリニアサーチ手段を有していることを特徴とする。

【００２０】上記サーチ手段は、木構造を用いて同時に

複数対のセルを評価していくロガリズミックサーチ手段を有していることを特徴とする。

【００２１】上記サーチ手段は、リニアサーチ手段とロガリズミックサーチ手段を結合したサーチ手段を有していることを特徴とする。

【００２２】上記タグ手段は、先頭から末尾までのキューを形成するように配置された一連のタグレジスタと、複数の比較器と、上記タグレジスタと比較器を接続する接続手段と、上記セルを対応するタグレジスタにより識別できるように記憶する記憶手段と、記憶されたセルのタグ値をタグレジスタに設定する設定手段とを備え、上記比較器は、タグレジスタのタグ値と１つ前の比較器からの出力値を比較して小さい方の値を決定するとともに、小さい方の値と宛先ビットを出力する出力手段と、小さい方の値と宛先ビットに基づいて、記憶されたセルを識別する識別手段とを備えたことを特徴とする。

【００２３】上記比較器は、一連のビットから構成された出力を持ち、この一連のビットの中に、宛先ビットが有意状態であり、同じ宛先ビットが有意状態であるキュー中の他のセルのタグ値よりも小さいタグ値を持つことを示す特定ビットを有しており、各比較器は、各タグ値とキューの末尾方向にあるタグの最小値とを比較することにより、宛先別に各セルのプライオリティを設定することを特徴とする。

【００２４】上記比較器は、２つのタグ値を比較する手段と、小さい方の値を出力する手段とを有し、１つのタグ値に対して１つの追加の入力ビットを備え、上記比較器の出力は、比較対象となるタグ値を示す追加の入力ビットにより条件付けられていることを特徴とする。

【００２５】上記追加の入力ビットは、宛先ビットが有意状態になっているセルのタグ値を特定するものであることを特徴とする。

【００２６】上記スイッチングシステムは、更に、最小のタグ値を持つセルが２以上存在する場合に、キューの先頭に近いセルを選択する手段を備えたことを特徴とする。

【００２７】上記タグ手段は、枝を経由してルートに連結された葉を持った階層的木構造を有し、各枝は、一対のタグレジスタと隣り合うタグレジスタの間に設けられたタグレジスタ用比較器と、２つのタグレジスタ用比較器からの出力を入力する追加の比較器を有しており、上記追加の比較器は、宛先ビットが有意状態になっており、小さい方のタグ値を持つタグレジスタに対応するセルを識別する手段を有していることを特徴とする。

【００２８】上記タグレジスタ用比較器は、小さい方のタグ値と２つの比較ビットを出力し、２つの比較ビットの内、一方の比較ビットは一方のセルが小さい方のタグ値を持っていることを示し、２つの比較ビットの内、他方の比較ビットは他方のセルが小さい方のタグ値を持っていることを示すものであり、上記選択手段は、小さい方のタグ値を持っているセルを識別するために、宛先ビットと比較ビットを入力する複数のＡＮＤゲートを有し、複数のＡＮＤゲートは、ルートにおいてただ１つのＡＮＤゲートが有意な出力を有するように構成されており、このＡＮＤゲートの有意な出力により出力リンクに出力されるセルを示すことを特徴とする。

【００２９】上記ＡＮＤゲートは、そのＡＮＤゲートに対応したセルが適切な宛先ビットを有しているかを判定するために用いられるものであり、上記タグレジスタ用比較器は、異なるＡＮＤゲートへ接続された出力ビットを有し、このタグレジスタ用比較器からの出力ビットは、対応するセルの宛先ビットとＡＮＤを取られ、その結果の出力が有意状態である場合は、対応するセルが宛先ビットが有意状態であり、最小のタグ値を持つセルであることを示すことを特徴とする。

【００３０】上記スイッチングシステムは、更に、２以上のセルが最小のタグ値を有する場合、先着のセルを識別する手段を有することを特徴とする。

【００３１】上記タグレジスタ用比較器は、先頭と末尾を持つキューを定義し、上記追加の比較器は、２入力から小さい値を出力するとともに、キューの先頭方向にある２入力の値がキューの末尾方向にある２入力の値以下である場合に、有意状態を出力することを特徴とする。

【００３２】上記スイッチングシステムは、更に、上記木構造のルート方向への上位層レベルにおいて、下位層レベルのどの枝が小さい方のタグ値を有しているか決定することにより小さい方のタグ値を有する枝を識別する識別手段を有し、上記識別手段は、各タグに対応してそのタグの宛先ビットとそのタグが最小のタグ値を有することを示すビットとのＡＮＤを取るためのＡＮＤゲートを有していることを特徴とする。

【００３３】

【作用】この発明のスイッチングシステムは、スケジュール手段がセルの出力をスケジュールする。スケジュール手段は、所定のアルゴリズムに基づいて、セルを順番に選択する。このように、スケジュール手段を備えることにより、セルの優先処理が行える柔軟なシステムを提供する。また、上記スケジュール手段は、いろいろな種類なアルゴリズムを採用することができる。そのアルゴリズムによって、優先度の高いセルをタイムリーに交換することができるとともに、優先度の低いセルを後から交換することができる。

【００３４】出力手段は、セルを出力する場合には、スイッチの内部で用いた宛先ビットのベクトルを除去して出力する。

【００３５】上記アルゴリズムとしては、ダイナミックなスケジューリングアルゴリズムやスタティックなスケジューリングアルゴリズムを用いることができるとともに、ラウンド・ロビン・アルゴリズムも用いることができる。或いは、これらのアルゴリズムの組み合わせでも

よい。

【００３６】スケジュール手段は、優先順位を示すタグ値を計算し、このタグ値を有するタグをセルに付加する。サーチ手段は、このタグ値と宛先ビットを用いて出力すべきセルをサーチする。結果として、選択手段は、ある出力リンクに対する宛先ビットが有意状態（オン）になっており、かつ、それらの中で最も小さいタグ値を持つセルを選択する。最も小さいタグ値を持つセルが２以上存在する場合には、先に到着したセルを選択する。

【００３７】キューイング手段は、複数のレジスタによりＦＩＦＯを構成している。この複数のレジスタからなるＦＩＦＯにより到着順を示すとともに、宛先ビットのベクトルとタグを保持するようにしている。

【００３８】サーチ手段は、複数の比較回路をリニアに配置し、各レジスタに記憶されたタグ値を順に比較することにより、最小のタグ値を持つセルを検出する。

【００３９】或いは、サーチ手段は、２個のレジスタに対して１つの比較回路を備え、２個のレジスタのタグ値の比較結果を更に追加の比較回路で比較することにより、ツリー形式を用いて、最小のタグ値を持つセルを選択する。このように、ツリー構造を用いた比較処理を行うことにより、キューに記憶されたエントリーの数の対数に比例する時間で最小のタグ値を持つセルを選択でき、前述したリニアにサーチする場合に比べて高速処理を行える。

【００４０】この発明のスイッチングシステムは、スケジュール手段が、セルの出力順番を予め設定されたスケジューリングアルゴリズムに基づいて決定する。スケジュール手段は、出力リンクに共通に設けられたキューを備え、そのキューに対して各セルの宛先ビットと優先度を示すタグを記憶させる。サーチ手段が、この宛先ビットとタグを参照することにより、出力リンクに出力するセルの順番を決定する。タグに記載されるタグ値は、ダイナミック、或いは、スタティックなアルゴリズムのいずれのアルゴリズムに基づくもので計算されたもので構わない。従って、このスイッチングシステムは、ユーザの要求に応じて柔軟なスケジューリング方式を採用することができる。

【００４１】上記サーチ手段は、リニアサーチ手段により、最小のタグ値を順番に捜していく。

【００４２】或いは、サーチ手段は、ロガリズミックサーチ手段を有し、木構造を用いて最小のタグ値を高速に検索する。

【００４３】或いは、サーチ手段は、リニアサーチ手段とロガリズミックサーチ手段を結合したサーチ手段により、最小のタグ値を検索する。

【００４４】上記タグ手段は、タグレジスタと比較器を備え、タグレジスタにタグ値を設定し、比較器がこのタグ値を順番に比較していくことにより、最小のタグ値を検出する。

【００４５】上記比較器は、特定の出力リンクに対する宛先ビットが有意状態（オン）であり、最小のタグ値を持つことを示す特定ビットを出力し、各比較器は、キューの末尾方向にある比較器からの出力と各比較器に対応したタグ値とを比較することにより、優先度を決定する。

【００４６】比較器は、２つのタグ値を比較して小さい方の値を出力する。また、比較器は、追加の入力ビットを備え、追加の入力ビットにより入力したタグを比較対象とするかどうかを決定する。

【００４７】上記追加の入力ビットは、宛先ビットが有意状態（オン）になっているセルを示すことにより、対応するタグ値が比較対象になるかどうかを特定するものである。

【００４８】もし、最小のタグ値を持つセルが２以上存在する場合には、先に到着したセルを選択する。

【００４９】前述したタグ手段の別な構成として、タグを階層的構造を持った比較器により比較するようにしても構わない。一対のタグレジスタに対して、比較器を１個設け、この比較器からの出力を更に追加の比較器で比較することにより、最小のタグ値を持つタグレジスタを検出する。このようにして、検索を高速に行う。

【００５０】比較器は、比較した２つのタグ値の内、いずれのほうが小さい値を持っているかを示す比較ビットを出力し、選択手段は、複数のＡＮＤゲートを有し、ＡＮＤゲートが比較ビットを入力するとともに、宛先ビットを入力することにより最終的に優先度の高いセルを１つだけ特定する。

【００５１】ＡＮＤゲートは、宛先ビットが有意状態（オン）になっているかどうかを判定するとともに、比較器からの比較ビットとのＡＮＤを取ることにより、最終的に宛先ビットが有意状態（オン）であり、最終のタグ値を持つセルを選択する。

【００５２】もし、２以上のセルが最終のタグ値を有する場合には、先に到着したセルを選択して出力する。

【００５３】タグ手段が前述したように、階層的木構造を用いて最小のタグ値を検索する場合には、タグレジスタ用比較器により先頭から末尾方向が存在するキューを定義し、追加の比較器により先頭方向にあるタグ値が末尾方向にあるタグ値よりも小さい場合に、その先頭方向にあるタグ値を持ったセルを選択する。

【００５４】本発明のスイッチングシステムは、更に、識別手段により階層的木構造を持つ場合に、どの枝が小さい方のタグを有しているかを識別し、その識別手段は、ＡＮＤゲートに基づいて、タグの宛先ビットとそのタグ値が最小のタグ値であることを示すビットとのＡＮＤを取ることにより、最終的に優先度の高いセルを選択する。このようなＡＮＤゲートを備えることにより、階層的木構造を取った場合に、葉からルートに対して宛先ビットを伝搬するため、各タグ毎に必要とされていたＡ

ＮＤゲートを省略することができる。結果として、ＡＮＤゲートの数を減少させることができ、回路構成を簡略することができる。

【００５５】

【実施例】

実施例１．各種のスケジューリングアルゴリズムを適合させるために、この発明に係るディジタル通信ネットワークスイッチは、改良されたスケジューリングシステムを備えている。その改良されたスケジューリングアルゴリズムによれば、各セルは、スイッチに到着する都度、タグ付けされ、その後、共通キューに記憶される。タグは、バイナリーの数値から成る。タグの数値は、トラヒックのクラスに対応するスケジューリングアルゴリズムや、セルが伝送されるバーチャルチャネル（ｖｉｒｔｕａｌ　ｃｈａｎｎｅｌ）の属性や、セル自身の特性を考慮して計算される。続いて、キューが宛先とタグにより出力リンク又は出力ポート毎に並列に検索される。複数のセルの宛先が同一である時、即ち、出力リンク又は出力ポートが同一である時は、その中で最小のタグ値を持つセルが選択される。これにより、スイッチは、常に、宛先毎に、スケジューリングアルゴリズムに従って、最初のセルを選択する。

【００５６】より一般的に言えば、この発明のスイッチングシステムは、アルゴリズムを記憶する手段を備えている。記憶されたアルゴリズムに従って、セルがスケジュールされる順番が決定される。また、アルゴリズムは、スイッチの動作に先立って、また、スイッチの動作中に変更可能であるので、ネットワークの幅広いトラヒック要求に対応することが可能となる。

【００５７】この発明のスイッチは、スタティックアルゴリズムとダイナミックアルゴリズムの両方をサポート可能である。これらのアルゴリズムには、以下のものを含む。ルー，ジャングによるレートモノトニックアルゴリズム、ジャングによる単純優先度アルゴリズム、クマーによるウェイテッドフェアキューイングアルゴリズム、ジャクソン，ルー，ホーン，ジェングによる最早デッドラインファーストアルゴリズム、ジャングによるバーチャルクロックアルゴリズム、カルマネクによるラウンド・ロビン及び階層的ラウンド・ロビン。以上のアルゴリズムは、本明細書で既に述べたものである。また、カングのフローコントロールドバーチャルチャネルアルゴリズムのような複雑な制御アルゴリズムもサポート可能である。

【００５８】この発明は、以下の３点により成り立っている。第１に、タグ値がいずれかの演算可能なファンクションに従って計算される点、第２に、タグ値がセルが伝送される順番を示している点、第３に、上に挙げたすべてのアルゴリズムに共通する特徴として、セルの順番が数値のタグで表現されると言う点である。検索を成立させる唯一の決定要因は、タグの持つ数値であり、タグの持つ意味は問われない。さらに、各セルのタグ値は、伝送中に計算されるので、スタティック・スケジューリング・アルゴリズムだけでなく、ダイナミック・スケジューリング・アルゴリズムもサポートできる。また、タグ毎に、十分な数のビットが用意されているので、スイッチバッファ中の全セルの順番を並び替えて、各ビットの値を書き変えることも可能である。

【００５９】この発明のポイントは、タグベースの検索を行う点にある。各セル（ＡＴＭセル）は、ネットワークスイッチに到着すると、バイナリーの数値を持つタグを付加される。典型的な実施例においては、８ビット、１６ビット、あるいはそれ以上の複数のビットを持つタグが用いられる。８ビットの場合には、２５６（２の８乗）通りのタグ値が設定できる。１６ビットの場合には、６５５３６（２の１６乗）通り、ｎビットの場合には、２のｎ乗のタグ値が設定できる。各セルのタグは、そのセルが伝送されるバーチャルチャネル、そのバーチャルチャネルのトラヒックのクラスのスケジューリングアルゴリズム、そのセル自身の特性等に関連する情報を持っている。そして、セル、タグ及び宛先情報が、伝送待ちのセルキューの末尾に入力される。この実施例においては、キューは、ＶＬＳＩ回路で構成されている。

【００６０】やがて、スイッチが、特定の宛先にセルを伝送しようとする時、キューに記憶されているすべてのセルを対象に出力リンク毎、或いは、出力ポート毎に並列に検索が行われ、その宛先に一致する最小のタグ値を持つセルが選択される。選択されたセルは、キューから消去され、伝送される。キューの中に同一のタグ値を持つ複数のセルが存在する場合には、もっとも早くキューに到着したセルが選択される。キュー検索部は、最小のタグだけでなくセルが伝送される出力ポートも識別する。

【００６１】このアーキテクチャーにより、スタティック・スケジューリング・アルゴリズム，ダイナミック・スケジューリング・アルゴリズムを問わず、前述した全てのアルゴリズムを含む幅広いクラスのスケジューリングアルゴリズムを実現することが可能となる。また、タグ値に範囲設定をすれば、１つのスイッチで同時に種々のトラヒックのクラスに対応する種々のスケージューリングアルゴリズムを実現できる。具体的には、タグ値を計算する際に、優先度の高いクラスのタグを数値の小さい範囲に割り当て、優先度の低いクラスのタグを数値の大きい範囲に割り当てる。そのような割り当てを行えば、数値を頼りに検索するシステムでは、優先度の高いクラスのセルが１つでもあれば、必ずそのセルが選択される。なぜならば、優先度の高いクラスのセルのタグは、優先度の低いクラスのセルのタグよりも、より小さい値を持つからである。

【００６２】図１を用いて説明する。ＡＴＭスイッチ１０は、複数の入力リンク１４を備えている。複数の入力

17

リンク１４は、複数の入力処理部１６に接続されている。複数の入力処理部１６からの出力は、それぞれキュー検索部１８とメモリ（セルバッファメモリ）２０に入力される。キュー検索部１８は、制御部２２により制御される。制御部２２は、また、入力処理部１６及び出力処理部２４も制御する。入力処理部１６及び出力処理部２４は、それぞれマイクロプロセッサとメモリを備えている。また、入力処理部１６と出力処理部２４は、制御部２２によりアルゴリズムを設定、或いは、再設定する手段を備えており、スイッチに到着するセルやスイッチから送り出されるセルのネットワークへの要求に従って、適切なアルゴリズムが用いられるように制御される。

【００６３】キュー検索部１８の出力は、それぞれ出力処理部２４に入力される。それに伴い、セルバッファメモリ２０の出力も出力処理部２４に入力される。２６に示すように、入力処理部１６の出力は、宛先情報とタグとバッファアドレスを含んでいる。これらは、キュー検索部１８に接続されている。また、２８に示すように、セルヘッダーとセル本体は、セルバッファメモリ２０に記憶される。

【００６４】次に、動作について説明する。セルは、入力リンク１４を介して入力処理部１６に到着する。入力処理部１６は、到着したセルをセル毎に処理する。通常、どのＡＴＭスイッチでも行われている内部処理に加えて、入力処理部１６はセル毎にタグ値を計算する。タグ値の計算は、そのセルのバーチャルチャネルに対応するスケジューリングアルゴリズムに従って行われる。タグ値は、また、バーチャルチャネルの状態やセルの到着時刻にも左右される。バーチャルチャネルとは、他の全てのデータストリームと区別される特定のデータストリームのエンドトゥーエンドのコネクションを指す。ＡＴＭネットワークにおいて、それは、セルヘッダー中のビットにより識別される。入力処理部１６は、また、バーチャルチャネルの状態情報やスイッチによりメンテナンスされる他の情報も計算し、更新する。

【００６５】上記計算後、タグと宛先情報とセルのバッファアドレスとが出力され、キュー検索部１８のキューに記憶される。また、セルヘッダーとセルデータは、セルバッファメモリ２０に出力される。その時点で、セルのアドレス、宛先及びタグはキューイングされ、セルヘッダーとセルデータは、セルバッファメモリに記憶される。こうして、入力処理部は、次のセルを受け付け、処理することが可能な状態になる。以上のことから、入力処理は、関連する入力リンク１４の１セルサイクルに割り当てられた時間内に完了しなければならない。１セルサイクルとは、入力リンクの帯域幅で１つのＡＴＭセルを受信するのに必要な時間のことである。

【００６６】キュー検索部１８は、キューを順に検索し、個々の出力処理部２４について対応する宛先ビット

18

を持つセルを選択する。出力処理部２４は、セルが伝送されるのに必要な計算があればその計算を行う。その後、ネットワークの他のノードやセルの最終宛先に向けて出力リンク３０にセルを伝送する。キュー検索部１８によって行われる検索は、全ての出力処理部に関連するスイッチ内の入力リンク１４及び出力リンク３０の内、最も遅いリンクの１セルサイクルに割り当てられた時間内に完了しなければならない。

【００６７】この発明のキュー検索システムの説明に先だって、図２を用いて既に知られている検索システムについて説明する。これは、図１に示すキュー検索部１８に相当する。図２に示す検索システムは、縦列接続することにより上位ステージ又は／及び下位ステージを備えることができ、拡張することも可能である。キュー検索部１８は、ＶＬＳＩを用いたＦＩＦＯ回路３１（ｓｅｌｆ－ｔｉｍｅｄ　ＦＩＦＯ）からなっている。キューエントリーは、ＦＩＦＯ回路３１の末尾から挿入される。ＦＩＦＯに未使用のエントリーがあれば、自動的にキューの先頭方向に前詰めにシフトされる。各キューエントリー３９は、レジスタに記憶され、宛先フィールド４１とアドレスフィールド３５と優先度ビット３７を持っている。宛先フィールド４１は、独立した一連のバイナリービットを持っており、各ビットは１か０かいずれかの値を持つ。この実施例では、１の値を有意状態（オン）としている。この値は、データがどの出力リンクに伝送されるかによって決定される。宛先フィールドを構成している各ビットを、宛先ビットを呼ぶ。スイッチの出力リンク３０と同数の宛先ビットがあり、各宛先ビットは、特定の出力リンクへのコネクションを示している。キューエントリーの宛先フィールドの対応するビットが１の時に、キューエントリーと出力リンクを結ぶコネクションが成立する。例えば、宛先フィールドの第３ビットがの１の時には、対応するデータは、第３の出力リンクに伝送されることを示している。

【００６８】上で述べたように、宛先フィールドの各ビットは、それぞれ出力リンクに対応しているので、キュー全体を縦の列に着目して先頭から末尾まで眺めてみると、キューに記憶している全てのセルと１つの出力リンクとの対応関係が分かる。即ち、各カラム（列）のビットが１か０かを調べると、セルが特定の出力リンクに伝送されるかどうかを確認することができる。マルチキャストセルの場合、即ち、１つのセルが複数の出力リンクに関連付けられている場合には、そのセルの宛先フィールドには、１の値を持ったビットが複数存在する。図２において、例えば、宛先フィールドの右端のカラムを見てみると、１の値を持ったセルが２つあることが分かる。この従来の検索システムにおいては、どのスケジューリングアルゴリズムを用いても、同一の出力リンクに向かう２つのセルがある場合に、その伝送の順番を決定できるようにするために、優先度ビット３７を設けてい

る。

【００６９】以上のように、キューエントリー３９の宛先フィールド４１の中に１の値を持つビットがあれば、そのキューエントリーのアドレスフィールド３５に示されたアドレスに対応して、セルバッファメモリ２０に記憶されているセルは、必ずそのビットに対応する出力リンクに伝送される。このことは、宛先ビットによってセルが１つの出力リンクに伝送されるか、或いは、複数の出力リンクに伝送されるかを識別できるということを示している。図２において、４３に示すキューエントリーのように、宛先フィールド４１の全てのビットが０である場合には、このエントリーは未使用であるとみなされる。また、図２には、検索回路５０とバス読み出し回路４５が示されている。

【００７０】次に、動作について説明する。制御部２２からキュー検索部１８に対して、「ある１つの出力リンク３０に伝送するセルを検索せよ」という命令が出されると、宛先フィールド４１の中でその出力リンクに対応するビットが特定される。次に、検索回路５０が特定されたビット位置に１がたっていて、かつ、キューの先頭に最も近いキューエントリー３９を選択する。次に、制御部２２は、選択されたキューエントリーのセルのアドレスフィールド３５をバス読み出し回路４５に入力する。バス読み出し回路４５において、選択されたセルのアドレスフィールド３５は、キュー検索部１８の出力４９となる。続いて、選択されたキューエントリーの対応する宛先ビットは、リセットされ、値は０となる。最後に、出力処理部２４は、アドレスを出力４９により受け取る。そのアドレスは、セルバッファメモリ２０から指定された出力リンク３０へ伝送するセルを選択するために用いられる。

【００７１】図３は、説明を簡単にするために、図２に示すキューエントリー３９が４つある場合を示している。また、図２において、宛先フィールド４１で示していたものの内、宛先ビットを４２として示している。また、遅延優先度フィルタ４８は、優先度ビット３７を入力する。キュー検索部１８は、出力リンク３０に対応するセルを選択するコマンドを受けると、宛先ビットの中からその出力リンクに対応するカラム４６を選択する。このカラムは、遅延優先度フィルタ４８に入力される。遅延優先度フィルタ４８には、既にキューの各エントリーに対応する２つの優先度を示す優先度ビットが入力されている。出力リンクが選択された時、宛先ビットの内でその出力リンクに対応するカラムの全てのセルに対応するビットが、遅延優先度フィルタ４８に入力される。遅延優先度フィルタ４８は、優先度ビットと宛先ビットを結合し、結合結果として０か１の値を持つ各セルに対応するビットを出力する。この時、出力された値が０であれば、そのセルは選択されないことを示している。また、ビットの値が１である時には、そのセルが選択される資格があるということを示している。このように、遅延優先度フィルタ４８は、２種類の優先度を用いて単純な優先度付けの機能を果たしている。この従来技術において、遅延優先度フィルタは、同じ宛先を持つ複数のセルの中から、いくつかの任意のセルを選択するために用いられている。より明確に言えば、遅延優先度フィルタは、高い優先度と低い優先度をつけることによって、高い優先度のセルが低い優先度のセルよりも、先に伝送されることを可能にしている。

【００７２】遅延優先度フィルタ４８から出力された一連のビットは、比較回路５０ａに入力される。比較回路５０ａは、各エントリーのビットをその両隣のビットと比較する。これは、図３に示されている論理和ゲート５２及び排他的論理和ゲート５４により行われる。その結果、出力ビットの内、１つのビットだけが１の値を持つことになる。それは、その１の値を持つビットに対応するエントリーが選ばれて伝送されることを示している。選択されたエントリーは、宛先ビットが１であるエントリーの内、最もキューの先頭に近いエントリーである。

【００７３】図３に示すアクセラレータ６０は、検索をスピードアップするために用いられる。その結果、１セルサイクル内で全ての出力リンクの検索を順に行うことが可能となる。アクセラレータ６０は、通常一般的に用いられている先読み回路であり、縦列接続される場合の下位ステージからの入力６２と上位ステージへの出力６４を備えている。下位ステージからの入力６２が０の時は、下位ステージにおいて、指定された出力リンクに対してエントリーが未選択であることを示している。下位のステージからの入力６２が１の時は、下位のステージにおいて、指定された出力リンクに対してエントリーが選択されたことを示している。従って、下位ステージからの入力が１の時は、比較回路からの出力ビット５６は、全て０になるように設計されている。上位ステージへの出力６４は、下位ステージ又は自ステージにおいて、エントリーが未選択の場合、０が出力される。下位ステージ又は自ステージにおいて、エントリーが既に選択された場合は、１が出力される。

【００７４】図４は、図３における従来の遅延優先度フィルタ４８を、この発明のリニアサーチ回路７０で置き換えたものである。この実施例においては、各キューエントリー３９は、タグレジスタ７２により拡張されている。タグレジスタ７２は、数値に対応する値を持っており、その数値の大きさは、優先度を設定するために用いられる。この発明においては、最も小さいタグ値を持つセルが最も高い優先度を与えられる。前述したように、タグの持つ値によって優先度が設定される。この実施例のリニアサーチ回路７０は、様々な選択機能を提供することを目的として備えられている。これは、前述した図３に示す遅延優先度フィルタ４８が単純化機能を目的としていた点と大きく異なっている。多様な選択機能を提

供することにより、複数のスケジューリングアルゴリズムの幅広い要求に柔軟に答えることができる。図４においても、説明を簡単にするために、数百のキューエントリーの中から、４つのエントリーを取り分けている。４つのエントリーに対して、それぞれタグレジスタ７２と比較器７４が対応している。各タグレジスタ７２は、図１に示す入力処理部１６で計算された、セルエントリーに対応するタグの値を示すバイナリーの数値を持っている。この数値が比較器７４において、キューの中の直前のタグの比較器の出力と比較される。比較結果により、２つの数値の内、どちらが小さいかが決定される。この順序付け処理の結果は、キューの中を順に次の比較器に引き継がれる。

【００７５】また、この順序付け処理の結果は、一連のビットとして出力ライン７６に出力される。これらの一連のビットの内、いずれかのビットに１がたっていれば、それは第１に、対応するキューエントリーの宛先ビットが１であることを示している。第２に、対応するキューエントリーのタグ値がキューの中で選択された宛先ビットが１である全てのキューエントリー中で、そのキューエントリーのタグ値が最も小さいということも示している。それ故、以下のことが分かる。１つは、リニアサーチ回路７０及び比較回路５０ａ及びアクセラレータ６０からなるリニアサーチ回路の出力ビット５６は、選択された宛先ビットが１にセットされている全てのキューエントリーの中から、１つだけを選択するということ、もう１つは、選択されたキューエントリーは、最小のタグ値を持つということである。また、更に、もしも２つ以上のキューエントリーが同一の最小のタグ値を持ち、同一の宛先ビットを持っている場合に、キューの先頭に最も近いエントリーが選択されるということになる。

【００７６】その結果、リニアサーチ回路７０により、スケジューリングアルゴリズムが適当であろう思って採用した方式に従って計算されたタグ値に基づいて、選択プロセスを独自に構築する機会が提供される。即ち、リニアサーチ回路７０は、タグ値が小さな値を持つセルを先に出力するように設計されていているのであるから、スケジューリングアルゴリズムが、先に出力したいセルのタグ値を小さな値にするということさえ守れば、どのような方式でタグ値を決定してもよいことになる。このようにして、スケジューリングアルゴリズムの独自構築が可能になる。

【００７７】特に、この発明においては、タグベースの検索を実現するために、従来の遅延優先度フィルタ４８を使用していない。遅延優先度フィルタ４８の代わりに、キューの各エントリーに対応するタグレジスタを持つ。また、各タグレジスタをキューの末尾に向かって他のタグと比較し、最小値を持つタグを選択する比較器７４を持つ。図５は、図４に示したリニアサーチ回路７０の詳細を示す図である。図５に示すように、以下の例において、ｋを各タグレジスタのビット数とする。また、Ｎをキューエントリーの総数とする。キューエントリーｉに対応するタグをタグｉとし、キューエントリーｉに対応する比較器を比較器ｉとする。タグｉのタグ値をタグ値T^iとする。キューエントリーｉに対応する宛先ビットを宛先ビットD^iとする。比較器ｉ＋１と比較器ｉの間を通っている一組のワイヤの出力値を出力値M^{i+1}とする。比較器ｉの出力を、出力値M^i，１ビットC^iとする。この１ビットC^iは、宛先ビットD^iがセットされているかどうかとともに、もし、セットされている時は、更に、タグ値T^iが出力値M^{i+1}以下であるかどうかを示す。

【００７８】キュー最後尾エントリーからの出力値M^Nは、宛先ビットD^Nに１がたっている場合には、タグ値T^Nの値にセットされる。また、宛先ビットD^Nが０の時は、2^k-1（全ビット＝１）という値になる。比較器ｉの一組のワイヤからの出力値M^iは、キューエントリーｉからキューの最終（キューエントリーＮ）までのエントリーの内、最小のタグ値を示す。また、ビットC^iは、宛先ビットD^iがセットされ、かつ、そのタグ値T^iがキューエントリーｉからＮまでのキュー内にあるタグ値の中で最小値の場合、１に設定される。出力されたビットC^iは、図２及び図３に示す従来例の回路の遅延優先度フィルタの出力を置き換える。検索回路５０は、単に、ビットC^iが１にセットされている最初のエントリーを選択する。即ち、最小のタグ値を持つエントリーを検索する。

【００７９】以下に、比較器の動作について、詳細に説明する。タグレジスタの各ビットを比較する比較論理回路を図６に示す。図６は、図５に示す比較器ｉのｊ番目の比較論理回路を示している。図６のT^i［ｊ］とラベル付けされたボックス１３０は、ｉ番目のキューエントリーのタグレジスタのｊ桁目（ｋ－１≧ｊ≧０）のビットを示している。M^{i+1}［ｊ］とラベル付けされたワイヤ１３２は、出力値M^{i+1}のｊ番目のビットであり、キュー最後尾方向にあるｉ＋１番目のキューエントリーの比較器ｉ＋１からくるものである。また、M^i［ｊ］とラベル付けされたワイヤ１３２は、出力値M^iのｊ番目のビットであり、このエントリーからキューの先頭に向かって出力されるものである。C^i［ｊ＋１］とラベル付けされたワイヤ１３６と、E^i［ｊ＋１］とラベル付けされたワイヤ１３８は、ｉ番目のエントリーの比較器ｉの（ｊ＋１）番目の比較論理回路からくるキャリービットである。また、C^i［ｊ］とラベル付けされたワイヤ１４０とE^i［ｊ］とラベル付けされたワイヤは、ｉ番目のエントリーの比較器ｉのｊ番目の比較論理回路から次の下位ビットの比較論理回路へ、即ち、（ｊ－１）番目の比較論理回路へ出力されるキャリービットである。

23

【0080】2つのキャリービットとワイヤからの出力値の意味は、以下のように定義される。

$C^i[j]=1$
if and only if $D^i=1$
and $T^i[(k-1)\cdots j]$
$\leqq M^{i+1}[(k-1)\cdots j]$ (1)

$E^i[j]=1$
if and only if $D^i=1$
and $T^i[(k-1)\cdots j]$
$=M^{i+1}[(k-1)\cdots j]$ (2)

$M^i[(k-1)\cdots j]$
$=\min\{2^k-1,$
$T^m[(k-1)\cdots j]$
for $m=i,\cdots,N,$
such that $D^k=1\}$ (3)

ここで、N＝キューのエントリー総数、N≧i≧1，k－1≧j≧0である。また、$T^i[(k-1)\cdots j]$は、kビットで示されたタグ値T^iの上位k－jビットの値を意味する。

【0081】式（1）により、$C^i[j]=1$の時は、宛先ビットD^iが1であり、かつ、エントリーiのタグ値T^iの上位k－1ビットからjビットまでの値がエントリーi＋1からエントリーNまでの最小のタグ値M^{i+1}の上位k－1ビットからjビットまでの値以下であることを示している。$C^i[j]=0$の時は、そのエントリーiは、検索すべきエントリーでないことを示す。$C^i[0]$は、C^iとして比較回路に出力される。また、式（2）により、$E^i[j]=1$の時は、$C^i[j]=1$の時で、かつ、タグ値T^iと出力値M^{i+1}の上位k－1ビットからjビットまでの値が等しいことを示している。従って、$E^i[j]=1$の時は、必ず$C^i[j]=1$である。式（3）は、$M^i[(k-1)\cdots j]$がエントリーiからエントリーNまでの最小のタグ値の上位k－1ビットからjビットまでであることを示している。

【0082】図6に示す比較論理回路の機能は、図7に示す論理値表で定義される。図7において、5行目の$C^i[j+1]$と$E^i[j+1]$の値は、ありえない組み合わせである。図7を用いて、比較器が最小値を導き出すことを以下に証明する。まず、最初に、以下のように、初期設定をする。

$C^i[k]=E^i[k]=D^i$
$M^N[j]=T^N[j]$ or $(-D^N)$,
for $j=0,\cdots,k-1$

【0083】式（1）及び（2）は、全てのi及びj＝kについて成り立つ。同様に、式（3）もi＝N及び全てのj＝0，・・・，k－1について成り立つ。次に、帰納法で話を進めよう。0≦j＜k，1≦i＜Nの時のビット$T^i[j]$を考えてみる。上記全ての値及び図6

24

のi，j以外の値について、仮定を満足したとする。すると、この仮定により、

$C^i[j+1]=1$
if and only if $D^i=1$
and $T^i[(k-1)\cdots(j+1)]$
$\leqq M^{i+1}[(k-1)\cdots(j+1)]$ (4)

$E^i[j+1]=1$
if and only if $D^i=1$
and $T^i[(k-1)\cdots(j+1)]$
$=M^{i+1}[(k-1)\cdots(j+1)]$ (5)

$M^i[(k-1)\cdots(j+1)]$
$=\min\{2^k-1,$
$T^m[(k-1)\cdots(j+1)]$
for $m=i,\cdots,N,$
such that $D^k=1\}$ (6)

となる。また、図7に示す比較器のロジックテーブルの1行目～4行目の場合には、$C^i[j+1]=0$であるから、$D^i=0$であるか、または、

$T^i[(k-1)\cdots(j+1)]$
$>M^{i+1}[(k-1)\cdots(j+1)]$

であり、1行目～4行目の5列目に示すように、

$M^i[j]=M^{i+1}[j]$

となる。その結果、

$T^i[(k-1)\cdots j]$
$>M^{i+1}[(k-1)\cdots j]$

以上のように、式（1），（2）及び（3）は、成立する。5行目は、ありえない$C^i[j+1]$と$E^i[j+1]$の組み合わせを示している。6行目～9行目の場合には、$C^i[j+1]=1$であり、$E^i[j+1]=0$なので、$T^i[(k-1)\cdots(j+1)]$が最小値となり、

$T^i[(k-1)\cdots(j+1)$
$<M^{i+1}[(k-1)\cdots(j+1)]$ (7)

及び、6行目～9行目の5列目に示すように、

$M^i[j]=T^i[j]$

となる。その結果、以下の式が成立する。

$T^i[(k-1)\cdots j]$
$<M^{i+1}[(k-1)\cdots j]$

また、6行目～9行目の6列目と7列目に示すように、$C^i[j]$と$E^i[j]$の値も対応してそれぞれ1及び0にセットされる。更に、式（7）と帰納法の仮定により、以下の式が成立する。

$M^i[(k-1)\cdots(j+1)]$
$=T^i[(k-1)\cdots(j+1)]$

その結果、m＝i，・・・，N，$D^m=1$に対して、以下の式が真となる。

25

M^i ［(k－1)・・・j］
＝T^i ［(k－1)・・・j］
＝min {2^k －1, T^i ［(k－1)
・・・j］, M^{i+1} ［(k－1)
・・・j］}
＝min {2^k －1, T^n ［(k－1)
・・・j］ (8)

１０行目～１３行目の場合には、C^i ［j＋1］＝1、かつ、E^i ［j＋1］＝1なので、T^i ［(k－1)・・・(j＋1)］は、今までの最小値M^{i+1} ［(k－1)・・・(j＋1)］と同じ値を示している。従って、

T^i ［(k－1)・・・(j＋1)］
＝M^{i+1} ［(k－1)・・・(j＋1)］

となる。それ故、１０行目と１３行目に示すように、T^i ［j］＝M^{i+1} ［j］の時、

T^i ［(k－1)・・・j］
＝M^{i+1} ［(k－1)・・・j］

となり、１１行目に示すように、T^i ［j］＜M^{i+1} ［j］の時、

T^i ［(k－1)・・・j］
＜M^{i+1} ［(k－1)・・・j］

となり、１２行目に示すように、T^i ［j］＞M^{i+1} ［j］の時、

T^i ［(k－1)・・・j］
＝M^{i+1} ［(k－1)・・・j］

となる。これにより、式（1），（2）及び（3）が導かれる。以上のように、式（1），（2）及び（3）は、キューに記憶されている全てのセルのタグに適合する。

【００８４】その結果、エントリーiのタグ値T^iが最小の時、

C^i ［0］＝C^i
T^i ［(k－1)・・・0］
＝min {2^k －1, T^n ［(k－1)・・・0］
for m＝1, ・・・, N, such
that D^n ＝1}

となる。以上のように、キューに対して高速なリニアサーチが行われ、最小値を持つタグが選択される。

【００８５】この比較回路には、例えば、ワイヤ１３６と１３８のように、２本のキャリーラインが必要である。もし、キャリーラインが１本しかない場合には、図７に示した７行目と８行目、１１行目と１２行目のように、T^i ［j］とM^{i+1} ［j］が異なる場合と、6行目と9行目、１０行目と１３行目のように、T^i ［j］とM^{i+1} ［j］が同じ場合との差異を検出することができない。

【００８６】この実施例のように、各キューエントリーの間に使用されるの単純比較器をＶＬＳＩで構成しリニアサーチを行うには、１ビットにつき、およそ２５個の

26

トランジスタが必要である。そのため、タグが１６ビットで構成されている場合には、約４００個のトランジスタが必要となる。M^1 ［0］の値を出力するためには、宛先ビットからの情報が必要である。この情報は、N個全てのキューエントリーのｋビットのタグの各ビットを逐次、走査比較することによって得られる。この走査比較は、配列の左上から右下に向けて順次実行される。１つのキューエントリーに対する比較結果が出るまで待つことなく、次のキューエントリーの第１ビット目の走査比較を先に開始しても構わない。このように、各比較器の走査比較を可能な限り早めに実行することにより、回路内の最長パスの長さは、N＋ｋビットとなる。

【００８７】キュー検索部を0．８mのＣＭＯＳで構成した場合、単純比較器は、およそ１６ビットのタグ比較を5．５ナノセカンドで行う。このことから、キューに１６個（N＝１６）のエントリーがあり、そのｊ番目のビットを走査比較するには、ほぼ同じ時間がかかることが想定できる。例えば、１６ビット（ｋ＝１６）のタグを有する１２８エントリー（N＝１２８）を持つキューから最小値を持つエントリーを選択するには、以下の式から概略４９．５ナノセカンドかかることが計算できる。

Ｎビットの走査比較時間＋ｋビットの走査比較時間＝（１２８／１６）×5．５＋5．５＝４９．５

【００８８】上記４９．５ナノセカンドは非常に速いが、この値は、キューエントリー数の増加により大きくなってしまう。例えば、キューエントリーが２５６個であれば、１２８個の場合のほぼ倍の時間がかかる。

【００８９】以上のような、キューをリニアサーチする方法には、時間がかかるという課題が残っている。この課題を解決するために、図８に示すような二分木の形式を用いてもよい。この改良例においては、一組（２つ）のキューエントリーのタグの間に１つの比較器を備えている。この比較器からの出力は、次の上位レベルの比較器の片側に入力される。あるレベルの２台の比較器に対応して、必ずその上位レベルに１台の比較器が配置されている。そのため、キューエントリーされているデータ数N＝2^n の時、比較器の数はリニアサーチの場合と同じく、2^n －１となる。図８においては、比較木（比較に用いる木構造）は、３レベルの深さとなっている。

【００９０】図８は、この実施例のキューエントリーが８個ある場合の比較器の配置を示す図である。モジュール８０は、二分木形式で２つの比較器から構成されるモジュールである。１つのキューエントリーからのタグレジスタ８２と、隣接するキューエントリーからのタグレジスタ８４が比較器８６に接続される。比較器８６において、入力された２つのタグの値が比較され、より小さい方のタグが選択される。選択されたタグに対応するキューエントリーの宛先ビットが１にセットされる。もしも、入力された２つのタグの値が同一である場合には、

27

比較器８６はキューの先頭により近い方のタグを選択する。このような構成を取れば、キューに記憶されている全てのエントリーから最小の値を持つタグを選択するのに必要な時間は、キュー内に記憶されているエントリーの数のログリズムに比例する。例えば、キューのエントリーが１２８個であった場合には、最小値を持つタグを選択するのに要する時間は、図４に示すリニアサーチ回路を用いた場合に要する時間の１／２以下となる。

例１．エントリーの数＝８の時、$\log_2 8=3$

例２．エントリーの数＝１２８の時、$\log_2 128=7$

【００９１】比較器８６の出力及び隣接する比較器８８の出力は、上位レベルの比較器９０に入力される。比較器９０の出力は、同様に、比較器９２に入力される。以上のように、図８においては、比較器の配置が木構造をなしている。

【００９２】各比較器からは、２つの出力ビットＣがＣ´（図中、バーＣで示す）出力される。例えば、比較器８６から出力された出力ビット９４は、論理積ゲート９８に入力され、もう１つの出力ビット９６は、もう１つの論理積ゲート１００に入力され、最小値を持つタグの選択に使用される。比較器からの出力ビットは、論理積ゲート９８と１００で、それぞれ選択されたキューエントリーの対応する宛先ビットとＡＮＤ演算される。その結果、論理積ゲートからの出力ビットが１である場合には、以下の３点を示している。第１に、対応するキューエントリーの宛先ビットに１の値が入っているということ。第２に、その比較器に入力された複数のタグからなるサブツリーの中で、最小のタグ値であるということ。第３に、同一の最小タグ値を持つエントリーが２つ以上あった場合に、そのエントリーがそれらの中で最も早くキューイングされていること、即ち、キューの先頭に近いことを示している。このように、比較器と対応する論理積ゲートを木構造で配置したことにより、該当する宛先ビットに１がたっており、かつ、キューの中で最小のタグ値を持つ最もキューの先頭に近いキューエントリーを選択することができる。

【００９３】比較器の動作についてより詳しく説明する。木構造の葉にあたる比較器は、ｋビットからなる２つのタグレジスタの値を入力し、一組のｋビットからなる出力を生成する。出力されたｋビットには、２つの入力の内、より小さい方の値が示されている。また、木構造の葉にあたる比較器の２つの出力ビットＣとＣ´の内、出力ビットＣに着目してみると、比較器への２つの入力の内、キューの先頭側からの入力値が、キューの最後尾側からの入力値よりもより小さい時、或いは、同一の値である時のみ、１がたつことになる。出力ビットＣが１の時は、出力ビットＣ´は０となる。出力ビットＣが０の時、出力ビットＣ´は１となる。

【００９４】各タグのタグ値は、宛先ビットの値を逆転

28

させた値でＯＲ演算される。これは、処理対象となっていないタグレジスタが小さい値を持っている場合に、現在処理中の宛先に関するタグレジスタの比較結果に、悪影響を及ぼすことを防ぐことを目的としている。宛先ビットが０の場合は、逆転させた値が１となり、この１とタグレジスタの各ビットのＯＲ演算すると、全ビットが１となり、最大のタグ値を示すことになる。このようにして、宛先ビットがのエントリーのタグ値を無視することができる。木構造のそれぞれのレベルにおいて、比較器からの出力ｋビットは、図面では比較器の右側に示されている。次の上位レベルの比較器の一方に入力される。比較器から出力されたＣ及びＣ´は、直前の下位レベルで選択された出力ビットとＡＮＤ演算される。ここでＣ´は、Ｃの値を逆転させた値を示す。その結果、検索の結果を示す出力ビットが最上位レベルにおいて出力される。最終的に１にセットされた出力ビットは、１つだけになる。即ち、最小値のタグを持つキューエントリーに対応するビットであり、かつ、キューの先頭に最も近いキューエントリーに対応するビットである。宛先ビットが１つも一致しない場合には、出力ビットが１つもセットされない。

【００９５】Ｎをキューエントリー総数とすると、この二分木からなる比較器の最長パスは、$k*\log_2 N$となる。前述したリニアサーチ回路と同様の考え方を用いれば、１６ビットのタグを有する１２８個のキューエントリーから結果を選択するまでにかかる合計時間は、$5.5\times\log_2 128=5.5\times7=38.5$ナノセカンドとなる。この時間は、キューのデータ数が２倍になっても、僅か５．５ナノセカンド延長されるだけである。例えば、２５６個のキューエントリーから結果を選択するまでにかかる合計時間は、$5.5\times\log_2 256=5.5\times8=44.0$ナノセカンドとなる。

【００９６】図８に示した回路においては、$N*\log_2 N$個の論理積ゲートが、最小値のタグを持った出力を選択するのに必要であった。この数は、各比較器からの出力Ｃ及びＣ´を次の下位層レベルにフィードバックすることで、２＊Ｎ個まで減少させることが可能である。図９は、２つの層からなる変形例を示す図である。図９において、使用されている符号で図８と同一符号のものは相当部分である。今まで示してきた図の中では選択出力が右向きであったのに対して、この図においては左向きとなっている。この回路においては、図８に示した回路よりもより多くの時間が必要となる。なぜならば、Ｃ及びＣ´が選択出力に加えられなければならないからである。この余分にかかる時間は、$\log_2 N$個の論理積回路の時間である。そして、論理積ゲート９８の出力は、論理積ゲート１０２に接続され、論理積ゲート１００の出力は、論理積ゲート１０４にそれぞれ接続される。これらの論理積ゲートには、図に示すように、宛先ビット１０６及び１０８もまた入力される。

29

【００９７】上で述べた二分木検索方法は、スピードが速く、また、リニアサーチ方式と同じ数の比較器を使用するものである。次に、より処理速度が速い変形例について述べる。図１０は、前述した２つの方式を組み合わせた構成図である。図において、キューは、Ｌ個のエントリーからなる複数のグループに分割されている。それぞれのグループには、図４に示した形式のリニアサーチ回路１２０が備えられている。これらのグループは、並列に検索される。その結果、グループ毎にグループの中で最小値のタグを持つエントリーを選択する。各グループからの出力１２４は、追加のリニアサーチ回路１２６に接続され、そこで全体の最小値のタグを持つグループが選択される。リニアサーチ回路１２６からの出力は、ビットの集まりからなる。そのビットの集まりの中の１ビットだけが、１の値を持っている。即ち、そのビットは、全体の中で最小のタグを持つグループに対応している。これらのビットの集まりは、次に論理積ゲート１２８の配列に接続される。この論理積ゲート１２８は、リニアサーチ回路１２０に対応している。リニアサーチ回路１２０の各出力ビットは、リニアサーチ回路１２６の出力ビットとＡＮＤ演算される。その結果、最終的に最小値のタグを持ち、かつ、処理対象の宛先ビットに１がたっているキューエントリーが選択される。

【００９８】以下に、より詳細に説明する。この結合方式においては、Ｌ個のエントリーからなるグループが並列して比較され、並列してリニアサーチ方式で比較される。そして、その結果が上位層のレベルで更に比較される。従って、各グループ毎にそのグループ内の最小値のタグを持つｋビットの値を生成する。更に、Ｌ本の出力ラインを生成する。木構造の中の末端でないグループの出力ラインは、折り返され、次の下位層レベルのグループの出力ラインとＡＮＤ演算される。それ故、葉に当たるグループの出力ラインは、そのタグがそのグループ内でも最小であり、更に全ての上位層レベルのグループ中で最小である場合のみ１がセットされる。

【００９９】各グループの最長パスの長さは、ｋ＋Ｌである。キューエントリー総数をＮとすると、グループ数は、Ｎ／Ｌとなり、$\log_L N$個の階層数が存在する。この場合、回路全体が最終結果を算出するまでの時間は、｛１階層が結果を算出する時間＋出力時間｝×階層数となり、（ｋ＋Ｌ＋１）＊$\log_L N$に比例する。例えば、１６ビット（ｋ＝１６）の比較器が５．５ナノセカンド必要とする場合、１６個のエントリー（Ｌ＝１６）を持つ１つのグループは、およそｋビットの走査比較時間＝５．５＋５．５＝２＊５．５＝１１ナノセカンドを検索に要することになる。16^2個＝２５６個（Ｎ＝２５６，Ｌ＝１６，１６グループ）のエントリーのキューは、$\log_L N = \log_{16} 256 = 2$となり、この数字（１１ナノセカンド）の倍だけの時間を必要とする。これは、約２２ナノセカンドである。また、木構造

30

の中で用いられる比較器の総数は、2^nである。Ｌの値に関わらず、比較器の総数は、およそ2^nとなる。

【０１００】より早い処理スピードを達成するために、必要であれば、木構造をパイプライン化することも可能である。だが、その場合には、回路もより複雑化するという欠点も発生する。階層化された比較器の各階層レベルは、パイプラインにおけるステージに相当する。第１サイクルで第１の出力リンクに対応する一組の宛先ビットが選択され、第１階層レベルの全ての比較器から第１階層レベルの結果が形成される。形成された結果は、次の第２階層レベルに出力される。そして、その次の第２サイクルにおいて、第２の出力リンクに対応した新たな一組の宛先ビットが第１階層レベルの比較器で選択される。一方、第２階層レベルの比較器では、第２サイクルにおいて、第１階層レベルの比較器からの出力、即ち、直前のパイプラインステージからの出力が比較される。$\log_L N$個のサイクルが経過した後、第１の出力リンクに対応する選択出力ビットが結果として得られる。それ以降のサイクルにおいては、１サイクル毎に各出力リンクに対応する結果がその都度得られる。このようにして、１ＡＴＭセル時間よりも、はるかに短い時間で各宛先の選択出力が得られる。

【０１０１】スケジューリングアルゴリズムをインプリメントするという観点から見ると、このキューイングシステムにより提供されるタグベースの検索方法は、大変柔軟性に富むものである。そして、スタティックなスケジュール方法やダイナミックなスケジュール方法いずれにも幅広く対応して実現可能なものである。以下に具体的に述べる。

【０１０２】単純優先度アルゴリズム及びレートモノトニックスケジューリングアルゴリズムにおいては、タグ値はバーチャルチャネルにより静的に割り当てられる。即ち、チャネルが設定されると、その優先度も決定される。そして、そのチャネルに到着するセルは全て同一のタグ値を割り当てられる。より小さいタグは、優先度のより高い優先度に対応しているので、スケジュールの際に優遇される。タグベースの検索は、2^kレベルの優先度、或いは、レートリゾリューションをサポートする。

【０１０３】ほとんどのＡＴＭスイッチは、いくつかの種類の優先度スケジューリングをサポートしている。だが、サポートされる優先度レベルの数は、非常に小さい。具体的には、およそ２レベルか４レベル位である。今日では、同一のネットワーク環境において、多種多様なタスクを柔軟にスケジューリングするために、かなり多くの優先度レベルをサポートすることが必要であることが明らかになっている。特に、レートモノトニックアルゴリズムは、反復性のリアルタイムタスクをそのタスクの反復の度数によって分類する。より高い度数を持つタスクは、より高い優先度を割り当てられる。多数の種類のタスクをサポートするために充分なリゾリューショ

ンを持つことが望まれていた。

【０１０４】バーチャルクロックアルゴリズム及びウェイテッドフェアキューイングスケジューリングは、単純なダイナミック・スケジューリング・アルゴリズムである。セルがＡＴＭスイッチに到着すると、到着時間とバーチャルチャネルの状態に基づいて、タグが計算される。同一のバーチャルチャネルの各セルは、異なるタグを付加される。セルがネットワークからどのようなサービスを受けるかも、タグを計算する場合に用いられるアルゴリズム要素の１つである。

【０１０５】最早デッドラインファーストスケジューリングアルゴリズムは、全てのスケジューリングアルゴリズムの中で、最も一般的なものである。一連のリアルタイムタスクがいかなる方法によりスケジュールされようとも、それらのタスクは、最早デッドラインファーストアルゴリズムにより、スケジュールされることが可能であるということが証明されている。バーチャルクロックアルゴリズムやウェイテッドフェアキューイングアルゴリズムのように、最早デッドラインファーストアルゴリズムもダイナミック・スケジューリング・アルゴリズムである。このアルゴリズムによれば、タグ値は、各セルの到着時間及びバーチャルチャネルの状態に基づいて計算される。その結果、タグは、デッドライン値を持つ。

【０１０６】最後に、ラウンド・ロビン・アルゴリズムは、ハードウェアで実現するよりも、ソフトウェアで実現する方がより実現しやすいと考えられている典型的なアルゴリズムである。このアルゴリズムを、この発明のタグベースの検索で実現することが可能である。それは、そのアルゴリズムを他のダイナミック・スケジューリング・アルゴリズムと同様に取り扱うことにより可能となる。そうすることによって、全てのバーチャルチャネルに対して、１つのグローバルカウンタＲが保持される。グローバルカウンタＲの値は、処理されたラウンドを記憶する。また、各バーチャルチャネルｖ毎に、そのバーチャルチャネルの最新の受信セルのラウンド番号を示す値 r_v が保持される。新しいセルを受信すると、タグは以下のように設定される。

tag_{cell}＝ｎｅｗ＿r_v

＝ｍａｘ｛Ｒ，ｏｌｄ－r_v｝＋１

【０１０７】次に、セルは、このタグを付加されてキューに挿入される。セルをディスパッチ（発送）する時間になると、最小値のタグが選択される。ラウンド・ロビン・チャネルの選択されたセルのタグがグローバルカウンタＲの値よりもより大きい場合、グローバルカウンタＲの値は、タグの値で書き換えられ更新される。

【０１０８】全てのダイナミック・スケジューリング・アルゴリズムにおいて、タグは任意の実時間時計により増加する要素であり、時としてタグ値の計算結果がｋ－１ビットを超えることがある。その場合、キューの全てのエントリーのタグ値は、ｋが充分な桁数を持っていれば、タグレジスタの最上位ビット（ｋ番目）が１となる。それ以上のオーバフローを防ぐためには、単にこれらのビットの全てをリセットしてゼロにし、処理を続ければよい。これは、ロジカルクロック又はバーチャルクロックの時間を 2^{k-1} ユニット毎に、2^{k-1} の値だけ進めることに等しい。より詳しく述べると、キューに残っている全てのエントリーのタグ値が 2^{k-1} を超え、新しいタグの値もまた 2^{k-1} よりも大きい時、全てのキューエントリーのタグレジスタの最上位ビットをリセットすることは、キューエントリーの順番を替えずに、キューにその時点で存在するタグ値及び以降キューに入ってくるタグ値を 2^{k-1} 減少させる効果がある。

【０１０９】実際のネットワークにおいては、トラヒックが何らかのパラメータにより保証された配信を要求した場合には、リアルタイムスケジューリングアルゴリズムを使い、保証された配信を要求しないトラヒックについては、他の公平、かつ、最適と思われるスケジューリングアルゴリズムを使うことが必要である。これは、タグの最上位ビットを使うことで、サポート可能である。例えば、あるスイッチのバーチャルチャネルがＣ個のクラスに分割されているとする。また、各クラスは隣のクラスの優先度よりも、より高い優先度であるとする。この時、クラスはタグの上位 $\log_2$ Ｃビットにエンコードされる。最も優先度の高いクラスは、最小の値を持つ。その結果、この発明のタグベース検索システムは、常に最も高い優先度のクラスから、最小値のタグを持つキューエントリーを選択する。

【０１１０】以上のように、この発明のディジタル通信ネットワーク用スイッチは、汎用のキューイングシステムを利用する。このシステムは、該当するエントリーで最小のタグを持つエントリーをキューから検索するという考えに基づいている。このようなシステムでは、リアルタイムの保証を要求するアプリケーションや、オーディオデータ、ビデオデータのような連続性のあるメディアや、迅速な応答を要求するアプリケーションをサポートする幅広いクラスのスケジューリングアルゴリズムを実現することが可能である。更に、１つのスイッチで同時に多数のスケジューリングアルゴリズムを実現できる。それにより、様々な基準に従って、様々なクラスのトラヒックをサポートできる。この発明は、プログラムにより実現してもよい。つまり、各入出力ポートに処理速度の速いプログラム可能なマイクロプロセッサをそれぞれ備え、このマイクロプロセッサ上で実行されるソフトウェアで、ＡＴＭセルに割り当てられたタグの計算方法を実現してもよい。このシステムは、ＡＴＭネットワークで種々のリサーチ及びプロトタイプ環境を柔軟に実現する新しいレベルを築き上げる。また、システムをインプリメントする開発者をネットワークサービスに関する負荷から解放し、システムをインプリメントする開発者をアプリケーションの仕様実現に集中させることがで

33

きるというメリットも得られる。

【０１１１】以上のように、この実施例においては、ＦＩＦＯアルゴリズム及びシンプルな優先度付け方式を含む幅広いクラスのセル・スケジューリング・アルゴリズムを実現するために、タグベースの検索を行うＡＴＭネットワークについて説明した。このスイッチングシステムにおいては、各セルはネットワークスイッチに到着すると、バイナリーの数値を持つタグを付加される。各セルのタグは、そのセルが伝送されるバーチャルチャネル、そのバーチャルチャネルのトラヒックのクラスのスケジューリングアルゴリズム、そのセル自身の特性等に関連する情報を持っている。そして、セル、タグ及び宛先情報が、伝送待ちのセルキューの末尾に入力される。この実施例においては、キューは、ＶＬＳＩ回路で構成されている。

【０１１２】スイッチが、特定の宛先にセルを伝送しようとする時、キューに記憶されているすべてのセルを対象に並列に検索が行われ、その宛先に一致する最小のタグ値を持つセルが選択される。選択されたセルは、キューから消去される。伝送され、キューの中に同一のタグを持つ複数のセルが存在する場合には、もっとも早くキューに到着したセルが選択される。キュー検索部は、最小のタグだけでなくセルが伝送される出力ポートも識別する。以上のように、スイッチキューに記憶された各セルには、バイナリーの数値を持つタグが付加される。それにより、キューは、ある特定の宛先を持つセルの中で、最小値のタグを持つセルを速やかに検索することが可能になる。この構成を取れば、様々な特徴を持つネットワークのトラヒックをサポートするほとんど全ての公表されているスケジューリングアルゴリズムを実現できる。また、シビアなリアルタイムの要求にも応えられ、連続性のあるメディアの伝送も可能で、迅速なレスポンスも実現できる。更に、この構成を取れば、独自のスケジューリング方式とアルゴリズムを持つ複数のトラヒックのクラスを、１つのネットワーク内でサポートできる。

【０１１３】

【発明の効果】以上のように、この発明によれば、柔軟性に富んだスケジューリングを行うことができる。また、いろいろな種類のスケジューリングアルゴリズムを採用することができる。

【０１１４】また、この発明によれば、宛先ビットのベクトルは、除去された形でセルが出力されるので、スイッチングシステム内部で用いた宛先ビットのベクトルが外部に出力されることはなく、スイッチングシステムのインタフェースは、従来のスイッチングシステムと同様であり、従来のシステムと互換性のあるシステムを提供することができる。

【０１１５】また、この発明によれば、各種スケジューリングアルゴリズムを採用することができ、ユーザの要

34

求に応じた、或いは、各システムの要求に応じたスケジューリングを行うことができる。

【０１１６】また、この発明によれば、タグという情報を用いることにより優先度を持った処理を行うことができる。

【０１１７】また、この発明によれば、レジスタの内部に宛先ビットとタグを保持しているので、レジスタを参照することにより、スケジューリングを行うことができる。

【０１１８】また、この発明によれば、サーチ手段がリニアにタグをサーチしていくので、単純な方法により出力すべきセルを検出することができる。

【０１１９】また、この発明によれば、サーチ手段が階層的木構造を用いてタグをサーチするので、高速にセルを検出することができる。

【０１２０】また、この発明によれば、スケジュール手段がタグ手段と共通のキューとサーチ手段を備えることにより、優先度付けを行ったスイッチング処理を行うことができる。

【０１２１】また、この発明によれば、リニアサーチ手段を備えているので、タグを順番に比べることにより、優先度の高いセルを検出することができる。

【０１２２】また、この発明によれば、ロガリズミックサーチ手段を備えているので、優先度の高いセルを高速に検索することができる。

【０１２３】また、この発明によれば、リニアサーチとロガリズミックサーチを任意に結合することができるので、柔軟な検索を行える。

【０１２４】また、この発明によれば、タグレジスタと比較器を備えているので、タグの値を比較器により順番に比較することにより、優先度の高いセルを出力することができる。

【０１２５】また、この発明によれば、比較器の出力の中に最小のタグ値を示す特定ビットを設けているので、この特定ビットを利用することにより、最小値を持つタグを検出することができる。

【０１２６】また、この発明によれば、比較器は、各タグに対して追加の入力ビットを備えているので、この追加の入力ビットにより、比較対象となるタグであるかどうかを判定することができる。

【０１２７】また、この発明によれば、追加の入力ビットを備えているので、この追加の入力ビットに宛先ビットの情報を与えることにより、対応するタグが比較対象となるタグであるかどうかを判定することができる。

【０１２８】また、この発明によれば、最小のタグ値を持つセルが２以上存在する場合でも、先着したセルを先に出力する。従って、同一の優先度を持つセルが存在する場合には、セルの逆転現象は生じない。

【０１２９】また、この発明によれば、階層的木構造を用いてサーチするので、高速サーチを行うことができ

35

る。

【０１３０】また、この発明によれば、比較器からの比較ビットと宛先ビットをＡＮＤゲートによりチェックしているので、最終的にただ１つのＡＮＤゲートが有意な出力を有するように構成できる。このように、比較ビットとＡＮＤゲートの採用により、簡単な回路により、優先度を持ったスケジューリングを行える。

【０１３１】また、この発明によれば、ＡＮＤゲートにより宛先ビットの検査と最小のタグ値があるかどうかの検査が行われるので、最終的に宛先ビットがオンであり、最小のタグ値を持つセルを検出することができる。

【０１３２】また、この発明によれば、２以上のセルが最小のタグ値を有する場合に、先着のセルを選択するので、同じ優先度を持つセルの逆転が生じない。

【０１３３】また、この発明によれば、階層的木構造を持ったサーチを行う場合に、キューの末尾よりもキューの先頭方向にあるものを優先して出力するので、先入れ、先出し方式を保ちながら、優先度を持ったスケジューリングを行うことができる。

【０１３４】また、この発明によれば、ＡＮＤゲートの数を減少させた回路を用いてスケジューリング処理を行うことができる。

【図面の簡単な説明】

【図１】　この発明のＡＴＭスイッチの論理的構造を示すブロック図である。

【図２】　従来例のキュー検索システムを示すブロック図である。

36

【図３】　図２に示す従来例の遅延優先度フィルタを備えた宛先ビットと検索回路の詳細図である。

【図４】　この発明の実施例の計算されたタグ値を選択するリニアサーチ手段に４個のキューエントリーを描いた概要図である。

【図５】　この発明のリニアサーチ手段の詳細図である。

【図６】　この発明のリニアサーチ手段の一要素のブロック図である。

【図７】　この発明の実施例の比較器の論理値の定義図である。

【図８】　この発明の実施例のロガリズミックサーチ手段に８個のキューエントリーを描いた概要図である。

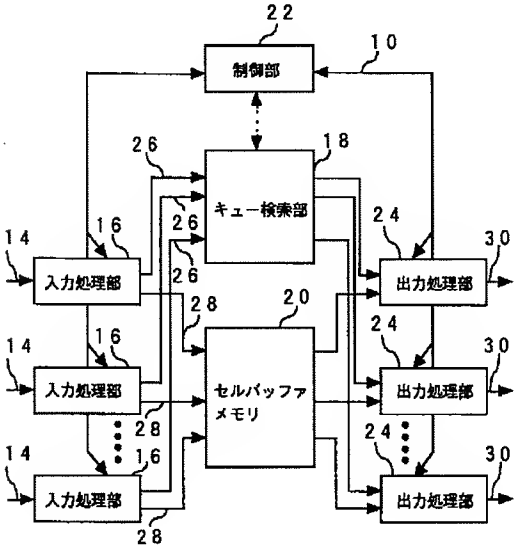
【図９】　この発明の実施例のロガリズミックサーチ手段の概要図である。

【図１０】　この発明の実施例のリニアサーチ手段とロガリズミックサーチ手段を結合させたブロック図である。

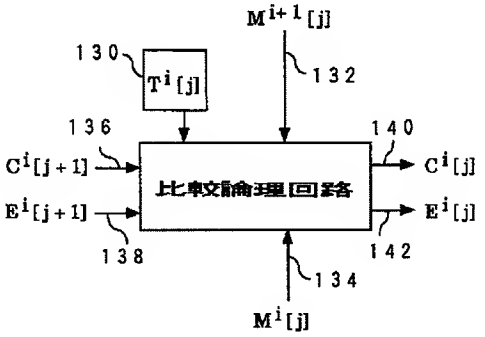
【符号の説明】

１０　ＡＴＭスイッチ、１４　入力リンク、１６　入力処理部、１８　キュー検索部、２０　セルバッファメモリ、２２　制御部、２４　出力処理部、３５アドレスフィールド、３７　優先度ビット、４１　宛先フィールド、４２　宛先ビット、５０ａ　比較回路、５６　出力ビット、７２，８２，８４　タグレジスタ、７４，８６，８８，９０，９２　比較器。

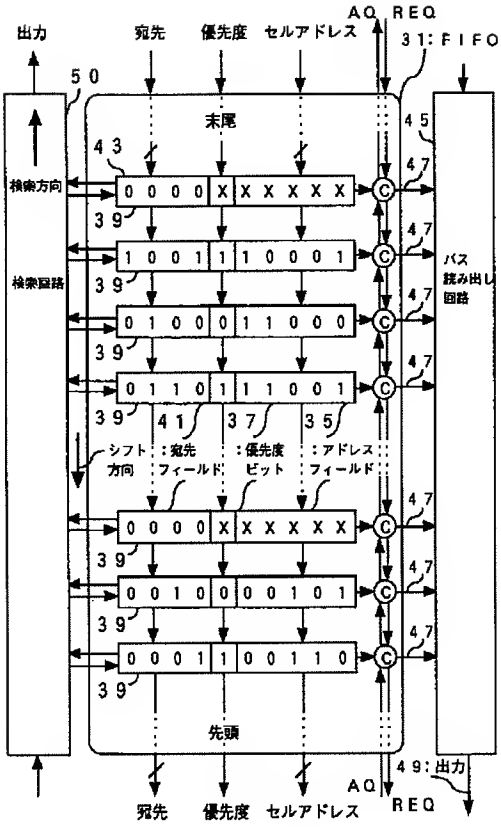
【図１】



【図６】



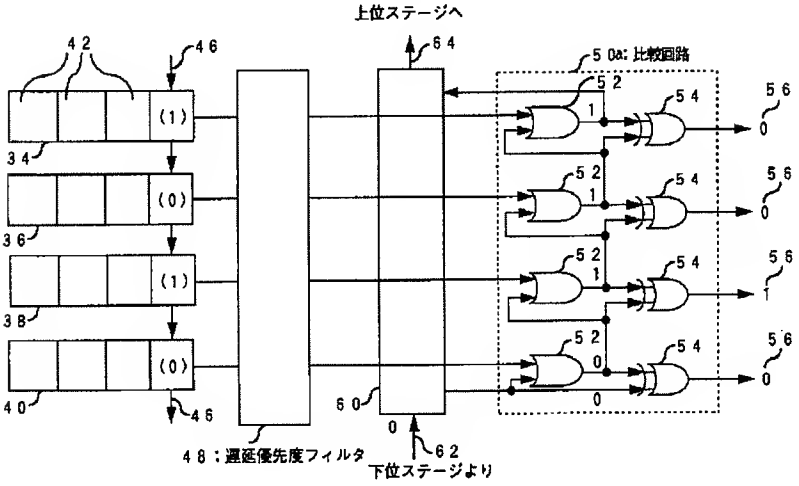
【図２】



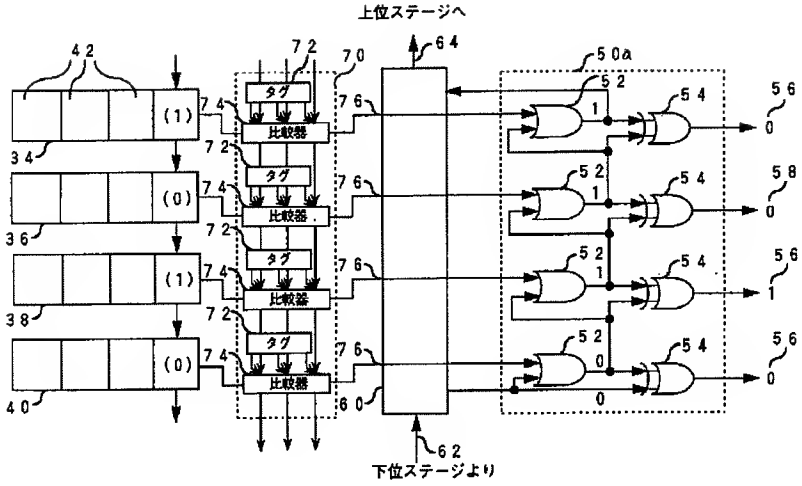
【図７】

	$C^i[j+1]$	$E^i[j+1]$	$T^i[j]$	$M^{i+1}[j]$	$M^i[j]$	$C^i[j]$	$E^i[j]$
1	0	0	0	0	0	0	0
2	0	0	0	1	1	0	0
3	0	0	1	0	0	0	0
4	0	0	1	1	1	0	0
5	0	1	X	X	X	X	X
6	1	0	0	0	0	1	0
7	1	0	0	1	0	1	0
8	1	0	1	0	1	1	0
9	1	0	1	1	1	1	0
10	1	1	0	0	0	1	1
11	1	1	0	1	0	1	0
12	1	1	1	0	0	0	0
13	1	1	1	1	1	1	1

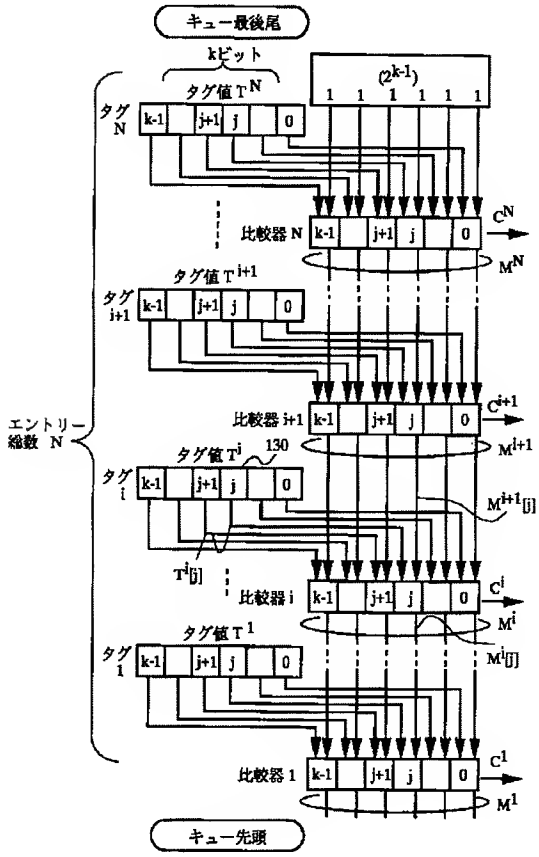
【図３】



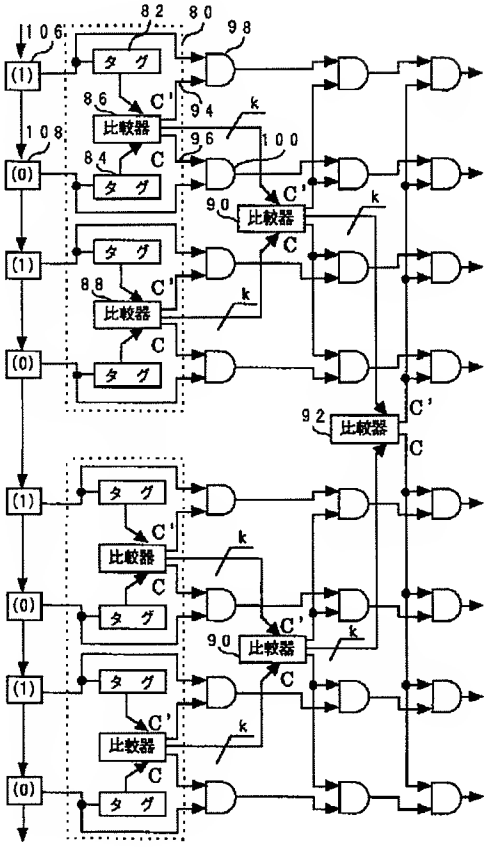
【図４】


上位ステージへ
42
72
70
64
50a
52
54
56
タグ
比較器
(1)
(0)
34
36
38
40
74
76
60
62
下位ステージより

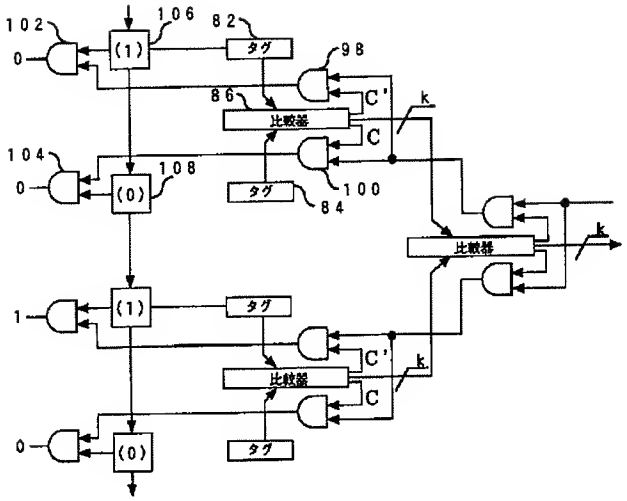
【図５】


キュー最後尾
kビット
タグ値 T^N
タグ N
比較器 N
C^N
M^N
タグ値 T^i+1
タグ i+1
比較器 i+1
C^i+1
M^i+1
エントリー総数 N
タグ値 T^i
130
タグ i
M^i+1[j]
T^i[j]
比較器 i
C^i
M^i
M^i[j]
タグ値 T^1
タグ 1
比較器 1
C^1
M^1
キュー先頭

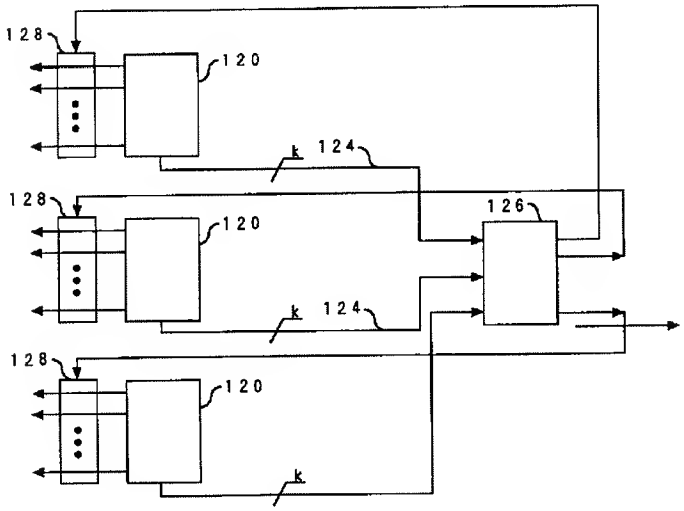
【図８】


82
80
98
106
タグ
(1)
86
C'
比較器
C
94
k
96
100
108
84
(0)
90
92
88

【図９】



【図１０】



フロントページの続き

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｈ０４Ｑ　3/52	１０１　Ｚ	9566－5G			
		9466－5K	Ｈ０４Ｌ　11/20	１０２　Ｚ	